



**INSTITUTIONEN FÖR FILOSOFI, LINGVISTIK  
VETENSKAPSTEORI**

# **METODENS STRINGENS OCH RESULTATETS RELEVANS:**

**RCTer och registerstudier som stöd i kliniskt arbete**

**Karin Svensson**

---

Uppsats/Examensarbete:	15 hp
Program och/eller kurs:	Magisterprogram i evidensbaserad
Nivå:	Avancerad nivå
Termin/år:	Vt 2016
Handledare:	Dick Kasperowski
Examinator:	Ingemar Bohlin

Uppsats/Examensarbete: 15 hp  
Program och/eller kurs: Magisterprogram i evidensbaserad  
Nivå: Avancerad nivå  
Termin/år: Vt 2016  
Handledare: Dick Kasperowski  
Examinator: Ingemar Bohlin  
Nyckelord: RCT, registerstudie, relevans, stringens, RRCT, medicin

## Abstract

---

Syftet med denna explorativa uppsats är att bidra till en ökad förståelse för vilken typ av kunskap om åtgärders effekter som RCTer respektive registerstudier kan bidra med vid ett kliniskt beslut. Utgångspunkt tas i begreppsparet stringens och relevans som är den framträdande konceptualiseringen av RCTer (stringens) och registerstudier (relevans). Induktionsproblemet är det vetenskapsteoretiska problem som ligger till grund för uppsatsen.

Litteratursökning har gjorts för att identifiera relevant litteratur om RCTer respektive registerstudier. Sökning har dessutom genomförts för att identifiera enskilda studier (RCTer och registerstudier) vars fokus är på åtgärder (både förebyggande och behandlande) vid diabetes typ 2. Vid granskning av litteratur och studier ligger fokus på design av patient och resultat.

Vid litteraturgranskningen framkommer att både RCTer och registerstudier har sitt ursprung i ett behov av kunskap om åtgärders effekter. Registerstudier och RCTer är lika såtillvida att de presenterar medelvärdesspatienter, samtidigt framträder spänningen mellan stringens och relevans. RCTens stringenta metod leder till ett trovärdigt resultat, men medelvärdesspatient i RCTen beskrivs i termer av orealistisk följsamhet, yngre, till större delen en man samt friskare än patienter i den kliniska praxisen. Det leder till att resultatets relevans för den kliniska praxisen ifrågasätts till skillnad från registerstudier. I registerstudier beskrivs medelvärdesspatienten ofta vara en spegling av den kliniska verkligheten, samtidigt som resultatets trovärdighet ifrågasätts. För att stärka registerstudiens trovärdighet lyfts olika faktorer fram.

Olika sätt att förstå begreppen relevans och stringens beskrivs i uppsatsen, vilket kan användas för att tydliggöra vilken typ av kunskap som respektive studieform kan bidra med vid ett kliniskt beslut. I uppsatsen beskrivs även hur RCTer och registerstudier kan ses hantera induktionsproblemet.

# Innehållsförteckning

1 Prolog .....	3
2 Teori och Begrepp .....	5
2.1 Induktionsproblemet.....	5
2.2 RCTer och registerstudier .....	6
2.3 Stringens och relevans.....	7
3 Syfte och Frågeställningar.....	9
4 Metod och Material .....	10
5 Resultat.....	12
5.1 Historisk utveckling .....	12
Jordbruksforskning i skapandet av stringens.....	12
Klinisk variation sätter strålkastarljuset på registerstudier.....	13
Valet mellan metodens stringens och resultatets relevans.....	14
5.2 Debatten fortsätter under det nya millenniet .....	15
Randomiseringens styrkor och svagheter .....	16
Effektstorlek: observationsstudiers möjlighet till övertag.....	17
Kvalitetsgranskningens betydelse vid bedömning av studiernas trovärdighet .....	18
Relevans genom överförbarhet.....	19
Går det att välja både stringens och relevans?.....	20
Registerbaserad RCT.....	20
Cirkulär evidensmodell .....	22
5.3 Design av patient och resultat i enskilda studier .....	23
Hur väljs patienter ut? .....	23
Hur skildras patienter?.....	26
Design av resultat .....	27
6 Diskussion .....	30
<b>6.1 Begreppsparet stringens-relevans.....</b>	<b>30</b>
<b>6.2 Design av patient och resultat .....</b>	<b>31</b>
<b>6.3 Olika aspekter av begreppet relevans.....</b>	<b>32</b>
<b>6.4 Olika sätt att se på stringens .....</b>	<b>33</b>
<b>6.5 Induktionsproblemet.....</b>	<b>33</b>
7 Epilog .....	35
Referenslista.....	37

# 1 Prolog

Jag arbetar som läkare inom primärvården, där jag bl.a. ingår i vårdcentralens diabetesteam. Nationella diabetesregistret (NDR) används rutinmässigt i mitt arbete. Data både rapporteras in till NDR och registret används för att följa upp vårdcentralens arbete. Det finns även tankegångar om att längre fram genomföra en registerstudie. Mot bakgrund av dessa rutiner och planer är det relevant att undersöka hur registerstudier kan informera den kliniska praktiken. Sedan länge har randomiserade kontrollerade studier (RCTer) här fyllt en viktig funktion. Den här uppsatsens tema är i vilken utsträckning registerstudier, med sin omtalade relevans i jämförelse med RCTers stringens kan utgöra stöd för den kliniska praktiken.

Uppsatsen tar sin början i en verklighetstrogen, men fiktiv dag på vårdcentralen. Första patienten för dagen är Pernilla som kommer för sin årliga diabeteskontroll. Pernilla tar dagliga promenader och äter hälsosamt under veckodagarna. Trots detta är jag fortfarande lite för kort för min vikt, säger Pernilla och skrattar. Hon använder redan två olika blodsockersänkande läkemedel. Trots det lyser blodsockervärdena röda i labblistan. Blodsockret ligger alldeles för högt. Med dessa blodsockervärden finns det risk att hon på sikt utvecklar njursvikt, nedsatt syn eller nervsmärtor. Lite senare under dagen träffar jag Tibor som även han har kommit för uppföljning av sin diabetes. Han är en tystlåten äldre herre. Innan besöket har jag snabbt hunnit skumma igenom hans journal. I journalen står det att han har haft tre tidigare hjärtinfarkter, dessutom följs han nu på medicinmottagningen då hans njurar sviktar. Under besöket framkommer att Tibor p.g.a. artros i knäna inte längre tar sina dagliga promenader. Han tackar däremot aldrig nej till en kaka eller bakelse när så bjuds. Tibor har i likhet med Pernilla, för högt blodsockervärde. Han har tidigare provat en blodsockersänkande tablett, men fick besvär med magen och klarade inte av att fortsätta. Vid bägge dessa verklighetstroga men fiktiva patientbesök uppstår frågan om det är aktuellt att sätta in något av de nyare läkemedlen som sänker blodsockret.

I Patientsäkerhetslagen (2010:659) står det att man som läkare ska arbeta "...i överensstämmelse med vetenskap och beprövad erfarenhet." För Tibor och Pernilla handlar det om relativt nya läkemedel och därför är den beprövade erfarenheten ännu inte så stor. Att arbeta i överensstämmelse med vetenskap kan vid en första anblick te sig som en självklar utgångspunkt, vilket också är i enlighet med grundtanken inom evidensbaserad medicin (EBM). SBU (Statens beredning för medicinsk utvärdering, 2014b s. 7) skriver att evidensbaserad medicin "...är ett förhållningssätt där man hela tiden kritiskt bedömer om vården vilar på bästa tillgängliga vetenskapliga grund." Vad innebär då bästa vetenskapliga grund? Inom medicinsk forskning beskrivs RCT (randomized controlled trial/randomiserad kontrollerad undersökning) oftast som golden standard (Bohlin & Sager 2011, s. 14). RCTen kännetecknas av att deltagare slumpmässigt fördelas mellan kontrollgrupp och interventionsgrupp.

Att de nyare läkemedlen sänker blodsockret framkommer tydligt i flera olika RCTer och meta-analyser (Socialstyrelsen 2015, ss 102-112). Detta faktum är inget som generellt ifrågasätts, däremot finns funderingar på om de effekter som fås i RCTer även kan fås för de patienter jag möter? De tillhör en värld som inte begränsas av olika inklusions- och exklusionskriterier. De kan ha sjukdomar eller vara av en ålder som medför att de inte hade inkluderats i RCTerna. Långtidsanvändning av dessa läkemedelsgrupper är dessutom inte lika välkänt som för de mer beprövade diabetesläkemedlen. Finns det några risker på lång sikt vid användandet av dessa läkemedel? För att bredda kunskapen om vilken effekt läkemedlen har i "real-world" lyfter man fram registerstudiers potential (Lee, Dekoven, Bouchard, Massoudi & Langer 2014, s. 820; Li, Chitnis, Hammer & Langer 2014, s 580.). På hemsidan för Nationella kvalitetsregister (SKL, <http://www.kvalitetsregister.se/forbattravarden.373.html>) står det att registerstudier kan bidra till tydligare riktlinjer och förbättrade rutiner. Det poängteras även att under senaste tiden har det kommit tunga vetenskapliga resultat baserade på data från Nationella Diabetesregistret (SKL,

<http://www.kvalitetsregister.se/forbattravarden/godaexempel/godaexempel/storaframgangarfordiabetes registretbaraborjan.1405.htm>).

Pocock och Elbourne (2000, s. 1907) skriver att observationsstudiers plats i utvärdering av behandling är ett långvarigt och omtvistat ämne. Förekomsten av olika uppfattningar om registerstudier respektive RCTers betydelse synliggörs även av Bohlin (2011, ss. 52-53) när han skriver:

Ingenting tyder på att den konflikt beträffande värdet av registerstudier respektive RCT:er som under tidigt 90-tal gav upphov till en livlig debatt, [...] skulle vara mindre djup i dag.

Så när beslutet ska tas ifall Pernilla och Tibor ska bli ordinerad något av de nyare läkemedlen, vilken hjälp kan jag få från RCTer respektive registerstudier? Vilken betydelse ska jag tillmäta resultat från dessa olika typer av studieform? Kan det var såsom Sager och Bohlin (2011, ss. 218-219) lyfter fram - att det handlar om en avvägning mellan metodens stringens och resultatets relevans? RCTerna baseras på en stringent metod, samtidigt som resultatens relevans ifrågasätts. Detta till skillnad från registerstudier som istället ses som en spegling av verkligheten, däremot är frågan vilken trovärdighet resultatet från registerstudier kan tillmätas, då registerstudier saknar RCTernas metodologiska stringens.

## 2 Teori och Begrepp

I denna del kommer jag först presentera det klassiska vetenskapsteoretiska problem som ligger till grund för uppsatsen. Innan en övergripande presentation av begreppsparet stringens-relevans ges, så beskrivs viktiga begrepp kopplat till de två typerna av studiedesign som är i fokus.

### 2.1 Induktionsproblemet

Inom vetenskapsteorin existerar ett antal klassiska problem som i grunden inte går att lösa utan som snarare måste hanteras på olika sätt. Det klassiska problemet som ligger till grund för denna uppsats är det s.k. induktionsproblemet. Induktionsproblemet säger att ett begränsat antal observationer aldrig kan producera en universell utsaga (Sismondo 2010, ss. 3, 30, 48). Två sätt att hantera induktionsproblemet på i medicinsk klinisk praktik är randomiserade kontrollerade undersökningar (RCTer) och registerstudier. Närhelst en vårdgivare möter en vårdtagare finns en potential att induktionsproblemet uppenbarar sig i den kliniska praktiken. Vem är denna patient i förhållande till de vetenskapliga resultat som skall vägleda olika terapier?

Chalmers (2013, s. 40-49) belyser frågan vad som utgör ett giltigt induktivt argument, alltså när det är legitimt att gå från ett ändligt antal observationer till en generell slutsats. Han lyfter upp tre förutsättningar som måste vara uppfyllda. De två första förutsättningarna är att antalet observationer måste vara stort samt att de ska upprepas under varierade förhållanden. Dessa två förutsättningar är dock problematiska. Ett stort antal observationer är ett vagt begrepp som inte tydliggör hur många det rör sig om. Dessutom kan det finnas situationer när det inte är adekvat att begära ett stort antal händelser (vilket Chalmers exemplifierar med användandet av kärnvapen och dess konsekvenser). Det är även oklart vad en signifikant variation av olika förhållanden innebär. Det tredje antagandet som Chalmers beskriver är att inga av de accepterade observationerna ska vara i konflikt med den generella slutsatsen. Det ses som problematisk av Chalmers (2013, s. 45) som skriver "...little scientific knowledge would survive the demand that there be no exceptions.". Sismondo (2010, s. 3) belyser ett annat dilemma kopplat till induktion, nämligen hur man kan veta om fall  $n+1$  är tillräckligt likt de  $n$  antal fall som ligger till grund för mönstret som man vill överföra till  $n+1$ . Det finns inte ett väldefinierat koncept för vad *sameness* är, vilket innebär att det finns flera olika tankegångar om vad det innebär att fallet  $n+1$  är samma som  $n$ .

Chalmers (2013, s. 49) konstaterar att det är svårt att specificera vad som utgör ett giltigt induktivt argument. Detta till skillnad från ett deduktivt argument som noggrant kan specificeras. Ett giltigt deduktivt argument innebär att om premisserna är sanna så är även slutsatsen sann. Däremot går det inte att enbart utifrån logik avgöra om premisserna är sanna, vilket Chalmers (s. 40) belyser med ett argument om katter. Chalmers argument består av två premisser, vilka är att alla katter har fem ben respektive att han har en katt. Slutsatsen blir således att hans katt har fem ben. Argumentet är logiskt giltigt, men inte sant.

Karl Popper var en av de som försökte kringgå problemen förknippade med induktion genom att via deduktion synliggöra en teoris konsekvenser så att dessa sedan kan testas (Chalmers 2013, ss. 54-57, 77, 80). Enligt Popper är nämligen en vetenskaplig teori falsifierbar. Premisser i giltiga deduktiva argument testas och klarar de inte testet förkastas teorin, samtidigt kan man utifrån denna ansats aldrig komma fram till att en teori är sann. Syftet med vetenskap beskrivs istället vara att försöka falsifiera teorier och när så sker ersätta dem med bättre teorier. Chalmers (2013, ss. 81) lyfter dock upp att falsifieringsansatsen inte är problemfri då det kan vara svårt att fastställa huruvida en teori är falsifierad i de situationer som observationerna motsäger teorin. Ett problem som belyses genom Duhem-Quine-tesen, vilken både Sismondo (2010, s.5) och Chalmers (ss. 82-83) beskriver. Duhem-Quine-tesen innebär att en teori aldrig kan testas isolerat, eftersom teorin utgör en del av ett nätverk som består av olika antaganden och rådande förutsättningar. Om en teoris förutsägelser är fel, innebär det inte att det måste vara teorin som är fel. Chalmers skriver att detta leder till att en teori kan

skyddas från falsifiering. Eventuella motsägelser mellan observationer och teorin kan nämligen bortförklaras genom att säga att det inte är teorin som är fel utan att det är en annan del av det testade nätverket som ger upphov till motsägelsen. Chalmers (s. 85) beskriver även ett annat problem förknippat med falsifiering, nämligen att teorier som anses vara bland de bästa exemplen på vetenskapsteorier aldrig hade utvecklats. Både Bohrs atommodell och Newtons gravitationslag är exempel på teorier som aldrig hade utvecklats ifall man följt en strikt falsifieringsmetod där teorier förkastats ifall de motsägs av observationer. Chalmers (s. 94) konkluderar därför att varken induktivismen eller falsifikationismen redogör för vetenskap på ett sätt som är kompatibel med vetenskapen.

## 2.2 RCTer och registerstudier

Randomiserad kontrollerad undersökning (randomized controlled trial, RCT) är en prospektiv undersökning. Det innebär att man börjar samla in data i samband med att deltagarna tas med i undersökningen. Deltagarna fördelas slumpmässigt mellan interventionsgrupp(er) och kontrollgrupp. Den slumpmässiga fördelningen beskrivs som en förutsättning för att okända störfaktorer (confounders) ska fördelas lika mellan grupperna. Fas I-III studier är exempel på studier där deltagarna vanligtvis är randomiserade. Fas I-III är kliniska prövningar som görs innan ett läkemedelsföretag ansöker om att ett läkemedel ska godkännas av läkemedelsmyndigheter. Fas I studie inkluderar oftast friska deltagare, då detta är den första studien av läkemedlet som görs på människor. Studiens fokus är i första hand biverkningar. Fas II studie syftar till att fastställa läkemedlets dosnivåer samt vilka biverkningar som dessa doser är förknippade med. Fas III studie är en större studie där fokus är på såväl effekter som på säkerhet. Fas III studier "...avser att ge de avgörande bevisen för läkemedlets effekt vid den eller de sjukdomar det är avsett för..." (SBU 2016).

Registerstudie är en undersökning som till skillnad från RCTen är retrospektiv (eller kombinerat retrospektiv- och prospektiv). Att en studie är retrospektiv innebär att data anskaffas innan undersökningen påbörjats. En registerstudie kan grundas på data som finns i offentlig statistik eller i medicinska databaser. En registerstudie kan utföras antingen som kohortundersökning eller som fall-kontroll undersökning. Kohortundersökning är en studie som gäller en grupp av personer som har vissa egenskaper gemensamt. Kohorten kan t.ex. bestå av personer som behandlats för en viss typ av sjukdom såsom diabetes. Olika undergrupper i kohorten jämförs med varandra. Undergrupperna kan exempelvis skapas utifrån att personerna behandlas med olika typer av blodsockersänkande läkemedel. På detta sätt kan utfallet för de olika läkemedelsgrupperna jämföras med varandra. Kohortundersökning kan vara såväl framåtblickande som bakåtblickande. Fall-kontrollundersökning innebär att i efterhand studera sambandet mellan exposition och sjukdom, vilket exempelvis kan vara att studera sambandet mellan rökning och lungcancer. Fallen är de som har sjukdomen (i detta fall lungcancer) medan de i kontrollgruppen saknar sjukdomen. Dessa grupper jämförs sedan beträffande expositionen (huruvida de har rökt eller inte). Såväl kohortundersökning som fallkontrollundersökning ingår i begreppet observationsstudier eftersom de inte har en experimentell studiedesign (SBU 2016).

En tydlig skillnad mellan RCTer och observationsstudier är att i RCTer randomiseras studiedeltagarna, vilket inte görs i observationsstudierna. Som tidigare beskrivits syftar randomisering till att okända confounders ska fördelas lika mellan grupperna. Relevanta skillnader mellan interventionsgruppen och kontrollgruppen kallas confounders (Howick 2011, s. 35-36). Howick lyfter fram tre typiska egenskaper som karakteriserar en confounder. En confounder ska potentiellt kunna påverka utfallet samt att den ska vara ojämnt fördelad mellan grupperna. Dessutom ska faktorn inte vara relaterad till interventionen. Detta innebär att en confounder kan innebära en möjlig alternativ förklaring till resultatet i en studie. Ett resultatfel som uppstått p.g.a. procedurfel, effektbedömningsfel, eller annat mänskligt fel under en undersökning kallas systematisk fel eller bias (SBU 2016). Det finns flera olika typer av bias, där ibland selektionsbias och behandlingsbias. Selektionsbias är kopplat till hur studien har skött urval av deltagare och indelning i interventionsgrupp- och kontrollgrupp. Selektionsbias kan leda till att grupperna inte är jämförbara med avseende på risk- och skyddsfaktorer. Risk för

behandlingsbias finns ifall interventionsgrupp och kontrollgruppen får felaktig eller ofullständig behandling eller om det förekommer tillägg utanför studieprotokollet (SBU 2014a, 2:6-2:7). För att hemlighålla viktiga förhållanden under studien, kan såväl deltagare som de som genomför studien vara blindade/maskerade. När studiedeltagare är blindade har de inte fått information om de tillhör interventionsgruppen eller kontrollgruppen (SBU 2016).

Registerstudier är det svenska begrepp som ofta används som översättning för det engelska begreppet outcomes research. Outcomes research används för att benämna den forskningsinriktning där fokus är på medicinska interventioners utfall (outcomes) och data hämtas från olika register och databaser (Bohlin 2011, ss. 35-36). Lee, Craig och Weeks (2000 ss. 198-200) skriver att outcomes research initialt handlade om effectiveness, alltså effekten av en intervention applicerade på en bred population. De beskriver att begreppet nu breddats, så att det är många fler studier som anses vara outcomes research. För att avgöra om en studie ingår i paraplybegreppet outcomes research, skriver Lee et al att man måste väga samman forskningsområde, studiedesign, metod och datakällan. Lee et al lyfter fram att outcomes research inte längre är begränsat till registerstudier. Utfall såsom hälsorelaterad livskvalitet anses vanligen vara utfall för outcomes research även om de samlas in genom fas III undersökningar. Jag väljer därför att i denna uppsats använda det begrepp som artikelförfattarna själva använder och inte per automatik översätta begreppet outcomes research med registerstudier.

## 2.3 Stringens och relevans

Sager och Bohlin (2011, s. 207-208, 218-219) lyfter fram att det i all kunskapsbildning görs avvägningar mellan metodens stringens och forskningsresultatets relevans. En avvägning som kan ses som ett val mellan studiens interna validitet och dess externa validitet. Intern validitet beskriver resultatets tillförlitlighet medan extern validitet avser i vilken utsträckning resultatet är giltigt utanför studien (SBU 2016). Sager och Bohlin (2011) skriver att RCTen har utformats för att sortera bort störande faktorer och variation för att inte påverka utfallet. En stringent metod leder till att orsakssamband mellan intervention och effekt blir tydliga. Sager och Bohlin tar samtidigt upp att det inte är säkert att sambandet som finns i RCTen även gäller i verkliga situationer, där störfaktorer och variation är en del av kontexten. Detta till skillnad från registerstudier som ses som en spegling av verkligheten, då variation och störfaktorer inte tas bort. Det innebär samtidigt att det är svårt att avgöra de verkliga orsakerna till de effekter som uppmäts. Valet mellan intern validitet och extern validitet har även uppmärksamats av Cartwright (2007, s. 11):

This is an instance of the familiar trade-off between internal and external validity. RCTs have high internal validity but the formal methodology puts severe constraints on the assumptions a target population must meet to justify exporting a conclusion from the test population to the target.

Cartwright (ss. 11-12) skriver att deduktiva metoder, vilket den ideala RCTen är ett exempel på, kan *clinich* (fastslå) ett orsakssamband. Det har dock sitt pris. Den stringenta metoden medför att resultatets externa validitet begränsas. Dessutom skriver Cartwright att det bara är specifika former av konklusioner som kan fås. I engelskan finns begreppet *efficacy* som används just för att belysa att effekten som beskrivs har erhållits i undersökningar genomförda under kontrollerade, laboratoriemässiga förhållanden. När effekten av kliniska interventioner undersöks i real-world används istället begreppet *effectiveness* (Raskin & Maklan 1991, s. 185).

SBU:s relevansbedömning (2014b, ss. 41-42) som görs inom ramen för systematiska översikter, synliggör att relevans inte alltid måste vara liktydigt med extern validitet. I SBU:s beskrivning av relevansbedömningen står att relevans inte ska förväxlas med studiens generaliserbarhet och den externa validiteten. Den specifika relevansbedömningen går istället ut på om studien uppfyller den systematiska översiktens inklusions- respektive exklusionskriterier. Relevansbedömningen handlar alltså i första hand om studien kan besvara ens fråga, inte om resultatet är överförbart från studiepopulationen till en annan population. Det kan samtidigt vara svårt att helt skilja på relevans och överförbarhet, då kriterierna kan vara utformade för att exkludera artiklar där man inte kan överföra



resultatet från studien till ens egen population. SBU:s relevansbedömning synliggör att det finns olika sätt att förstå relevansbegreppet. Detta är i samma anda som Knaapens artikel (2013, ss. 688-689). Knaapen belyser att relevans kan handla om vilka studier som är av betydelse för skapandet av en rekommendation. Relevans kan även ses som en bedömning av likheter mellan de experimentella förhållanden och förhållanden som råder där rekommendationen är tänkt att gälla. Syftet med denna bedömning är att undersöka om studieresultatet går att överföra till den kliniska verksamheten. Cartwright och Hardie (2012) synliggör i sin bok att det inte räcker med att veta att en intervention fungerat *där* i en RCT för att kunna säga att interventionen kommer fungera *här*. Man måste även kartlägga om interventionens stödfaktorer (de övriga faktorer som behövs för att interventionen ska vara verksam) som finns i RCTen även finns *här*. Man måste dessutom bedöma om interventionen kan utgöra samma orsaksroll *här* som den gjorde i RCTen.

### 3 Syfte och Frågeställningar

Uppsatsens syfte är att bidra till en ökad förståelse för vilken typ av kunskap om åtgärders effekter som RCTer respektive registerstudier kan bidra med vid ett kliniskt beslut. För att uppnå huvudsyftet består uppsatsen av två delar som bägge belyser design av resultat och patient i RCTer respektive registerstudier. Utgångspunkt tas i begreppsparet stringens och relevans som är den framträdande konceptualiseringen av RCTer (stringens) och registerstudier (relevans).

Uppsatsens första del fokuserar på RCTers respektive registerstudiers historiska bakgrund och utveckling, för att se hur design av patient respektive resultat beskrivs i litteraturen. Frågeställningar kopplade till denna del är:

- 1.1) Vilka styrkor och svagheter kopplas samman med RCTer respektive registerstudier?
- 1.2) Hur beskrivs RCTers respektive registerstudiers trovärdighet och användbarhet?

Uppsatsens andra del koncentrerar sig på enskilda RCTer respektive registerstudier, där fokus smalnas av och ligger specifikt på åtgärder vid diabetes typ 2. Här undersöks designen av patient och resultat i enskilda RCTer respektive registerstudier. Följande frågeställningar ställs i denna del:

- 2.1) Hur väljs patienter ut för att delta i studierna?
- 2.2) Hur stor är studiepopulationerna?
- 2.3) Hur beskrivs de inkluderade patienterna?
- 2.4) Vad studeras och vilka utfallsmått används?
- 2.5) Vilka typer av slutsatser drar artikelförfattarna utifrån studiernas resultat?

## 4 Metod och Material

Denna uppsats är av explorativ karaktär. Både litteratur som behandlar RCTer respektive registerstudier samt enskilda studier har granskats. Litteratur som inkluderades i denna uppsats studerades utifrån fokus på design av patient och resultat i RCTer respektive registerstudier samt kopplingen till begreppen relevans och stringens. Design av patient avser aspekter såsom hur patienter väljs ut för att ingå i studier, antal inkluderade patienter och hur patienterna beskrivs. Design av resultat avser exempelvis vad som studeras och vilka utfallsmått som används.

För att kunna besvara frågeställningarna i uppsatsens första del har jag under våren 2016 sökt efter relevant litteratur. Utgångspunkt för min litteratursökning har varit Bohlins kapitel *Evidensbaserat beslutsfattande i ett vetenskapsbaserat samhälle. Om evidensrörelsens ursprung, utbredning och gränser*. (2011), SBU:s metodhandbok (2014b) och Howicks bok *The Philosophy of Evidence-Based Medicine* (2011). För att identifiera ytterligare litteratur av intresse har jag studerat referenslistor i de ovannämnda texterna och i de texter som jag senare identifierade som relevanta. Jag har även sökt i web of science för att se vilka som senare citerat de artiklar jag har bedömt vara relevanta. Jag har även använt mig av kvalitetsregistrets hemsida. Jag har dessutom sökt efter artiklar i läkartidningens artikelarkiv på nätet. Söktermer som har använts var registerstudier, RCT, RRCT och observationsstudier. Sökningen av litteratur har skett kontinuerligt under arbetet med uppsatsen fram tills jag bedömde att den första frågan kunde besvaras. Målet med litteratursökningen har inte varit att genomföra en heltäckande sökning som inkluderar all tänkbar litteratur som publicerats av relevans för denna uppsats. Syftet har istället varit att försöka belysa registerstudier respektive RCTer ur olika vinklar, för att på detta sätt se hur designen av patienter och resultat beskrivits samt vilken betydelse det ses ha för studiernas användbarhet och trovärdighet. De båda studieformernas styrkor och svagheter, beskrivet av olika författare, identifierades och grupperades i olika teman. De inkluderade texterna har analyserats för att se vilka olika aspekter av relevans respektive stringens som jag kunde notera.

Anders Ekblom, professor i klinisk epidemiologi, lyfte vid Läkarsällskapets debatt om olika vetenskapliga metoder upp att Sverige har en fantastisk fördel vid genomförandet av registerstudier. Detta är kopplat till att vi har personnummer och därmed kan länka samman data från olika register (La Folie 2002, s. 3132). När uppsatsens andra frågeställning skulle besvaras har jag därför funnit det vara av intresse att jämföra svenska registerstudier med RCTer genomförda i Sverige. Jag har utgått från listan över registerstudier på NDRs hemsida (<https://www.ndr.nu/#/forskning>). Listan består av publicerade registerstudier som använt sig av NDR (Nationella Diabetesregistret). Vid genomgång av listan 16 april 2016 fanns 92 artiklar. Via titlar och vid behov även abstrakt har jag identifierat de artiklar som jag bedömde studerade effekter av olika åtgärder. Dessa artiklar har jag sedan rangordnat efter antal citeringar per år. För att få fram det totala antalet citeringar har jag använt mig av web of science. Valet att använda mig av web of science grundas på att den innehåller medicinska artiklar samt att artiklarna kan rangordnas efter totala antalet citeringar. För att identifiera RCTer har jag sökt i web of science med kombinationen "diabetes typ 2" AND (randomized OR RCT) AND (Sweden OR swedish). Sökningen gjordes 28 mars 2016 och gav då 103 träffar. Genom läsning av titel och vid behov även abstrakt identifierade jag de artiklar som baserades på RCTer och som åtminstone till viss del var genomförda i Sverige. Dessa artiklar har jag sedan rangordnat efter antal citeringar per år. Därefter har jag summariskt läst igenom cirka 10 artiklar av vardera studiedesign. Jag har sedan valt ut artiklar för att få en bredd i artikeltyp. Här har jag utgått från de olika typer av åtgärder (läkemedelsbehandling, kirurgiska ingrepp, livsstilsförändringar) som studierna undersöker. Jag har även valt att inte inkludera flera olika artiklar från samma studie. För att uppnå mättnad inkluderade jag fem registerstudier respektive fem RCTer för fördjupad granskning. Att inkludera ytterligare artiklar bedömde jag inte skulle bidra ytterligare till denna studie.

Sökning efter RCTer som gjordes i mars hade inte med "randomised" och därför gjorde jag en förnyad sökning 3 juni 2016 för att se om relevanta studier hade missats. Sökningen gjordes i web of science.

Förutom ovannämnda söksträng användes även en liknande söksträng men där randomized ersattes med randomised. Bägge dessa söksträngar gav 106 träffar var. När de bägge söksträngarnas resultat jämfördes med varandra var de 106 träffar identiska. Utifrån denna jämförelse bedömer jag därför att sökningen som gjordes i mars inte innebar att viktiga artiklar missades trots att söktermen "randomised" då inte användes.

Jag granskade de enskilda studierna utifrån följande områden: studiens syfte, urvalspopulation, antal inkluderade patienter, inklusions- respektive exklusionskriterier, patientkaraktäristika, utfallsmått, vilken typ av svar som studien ger samt författarnas egna tankar om studiens styrkor och svagheter. Denna granskning syftade till att se hur design av patient och resultat sker i enskilda studier. Vid analysen av de enskilda studierna letades både efter kopplingar och motsägelser till de beskrivningar av design av patient och resultat som har framkommit i uppsatsens första del. Precis som vid granskningen av den historiska och samtida litteraturen så sökte jag även i de enskilda studierna efter olika aspekter av begreppen relevans respektive stringens. Målet med granskningen av de enskilda RCTerna och registerstudierna syftade inte till att se ifall den ena typen av studieformat över- eller undervärderar en interventions effekter. Denna typ av undersökning är något som redan är gjort av flera olika författare, vilket vi kommer se längre fram i denna uppsats.

## 5 Resultat

I denna del kommer i tur och ordning svaren på frågeställningarna att presenteras. Vi börjar med en tillbakablick för att se hur patienter och resultat i RCTer respektive registerstudier har beskrivits (5.1), för att sedan närma oss ett mer nutida perspektiv (5.2). Därefter presenteras granskningen av de inkluderade RCTerna respektive registerstudierna (5.3).

### 5.1 Historisk utveckling

Bohlin (2011, s. 34, 49) beskriver hur både RCTer och registerstudier (genom resultatrörelsen) har bidragit till framväxten av evidensbaserad medicin. Den historiska bakgrunden och utvecklingen skiljer sig dock åt för respektive studiedesign. RCTer har sitt ursprung från det brittiska jordbruket. Resultatrörelsen (där registerstudier utgör den bärande idén) utvecklades inom den amerikanska hälso- och sjukvård under 1980-talet.

#### Jordbruksforskning i skapandet av stringens

Randomisering som en forskningsmetodologi utvecklades under 1930-talet inom jordbruksforskning av Ronald A. Fisher. Detta för att jämföra avkastningen av olika typer av säd. Genom att slumpmässigt fördela utsädet mellan olika delar av åkrarna, ville man säkerställa att uppmätta skillnader berodde på sädens egenskaper och inte p.g.a. variation av jordens sammansättning (Bohlin 2011, s. 39). Att använda sig av randomiserade, kontrollerade studier för att utvärdera läkemedelseffekt var ingen självklarhet under 1930-talet. I sina försök att övertyga läkare om behovet av att genomföra kontrollerade studier, skriver Bradford-Hill (1990, ss. 77-79) att han under 1930-talet inte vågade nämna ordet randomisering, eftersom han var rädd att det skulle skrämma bort läkare. Dr Philip Darcy Hart, som var aktiv inom behandling av tbc, såg också att det fanns ett behov av att genomföra kontrollerade undersökningar av läkemedel. Tbc var under 1940-talet en ledande dödsorsak bland unga vuxna. Guld hade använts under många år för att behandla tbc, vilket frustrerade Hart. Det var nämligen oklart om guldbehandlingen var effektiv. Hart var en av de som tillsammans med Bradford-Hill kom att genomföra studien som 1948 publicerades i *British Medical Journal* under titeln *Streptomycin treatment of pulmonary tuberculosis*. En studie som kom att ses som stilbildande inom klinisk forskning då det var en dubbelblindad, randomiserad kontrollerad undersökning (Bohlin 2011, s. 40). I artikeln från 1948 (*Streptomycin in Tuberculosis Trials Committee*, s. 769) framgår i introduktion varför det vid behandling av tbc inte går att förlita sig på enskilda fall eller fallserier för att bedöma läkemedels effektivitet:

The natural course of pulmonary tuberculosis is in fact so variable and unpredictable that evidence of improvement or cure following the use of a new drug in a few cases cannot be accepted as proof of the effect of that drug. [...] The history of chemotherapeutic trials in tuberculosis is filled with errors due to empirical evaluation of drugs (Hart, 1946); the exaggerated claims made for gold treatment, persisting over 15 years, provide a spectacular example. It had become obvious that, in future, conclusions regarding the clinical effect of a new chemotherapeutic agent in tuberculosis could be considered valid only if based on adequately controlled clinical trials.

Det var dock inte bara det stora behovet av ett effektivt läkemedel som ledde fram till den första randomiserade läkemedelsstudien. Bradford-Hill (1990) lyfter själv fram de knappa ekonomiska resurserna som rådde efter andra världskriget, som en möjlig bidragande faktor. Streptomycin som hade uppfunnits i USA 1944 var mycket dyrt. Därav hade man vid tidpunkten för genomförandet av RCTen bara råd med cirka 50-tal behandlingar. Det fanns alltså för få behandlingar sett till antalet som var i behov av dem. Bradford-Hill skriver att hade behandlingen funnits tillgänglig i större mängder hade studien troligen inte genomförts.

I skildringen av RCTens intåg i den medicinska världen framkommer en önskan om att komma åt det som en stringent metod ses utlova – nämligen en säkerhet om att de noterade effekterna beror på åtgärden som studeras. Det kan ses som ett behov sprunget från en klinisk verklighet som är så brokig

och varierande att det inte går att lita på enskilda observationer när man ska uttala sig om ett läkemedels effekter. Metodens stringens kan ses bidra till en ökad relevans sett ur ett patientnyttoperspektiv, då det varken för den enskilda individen eller för en större population ter sig vara relevant att ordineras ett läkemedel som varken medför chans till förbättring eller lindring vid ett dödligt tillstånd. Det går samtidigt att ana att de kliniska läkarna hade en farhåga om att den stringenta metoden hade sitt pris, detta då de beskrivs vara tveksam till införandet av kontrollerade studier.

Det skulle sedan dröja innan RCTen ansågs vara en allmänt accepterad metod inom medicin. Under 1960-talet sågs RCTen fortfarande som en underlighet (Evidence-Based Medicine Working Group (EBMWG) 1992, s. 2420), men förändringar skedde inom den kliniska forskningen och RCTens status höjdes. EBMWG (s. 2420) skriver 1992 att praktiskt taget inget nytt läkemedel kan bli en del av klinisk praxis om inte dess *efficacy* visats i kliniska undersökningar. RCTers position är fortsatt starka än idag. Bohlin (2011, ss. 38-40) skriver att idag ses RCTer generellt sett som den typ av studieform som ger de mest trovärdiga resultaten när behandlingseffekter ska studeras. McMastergruppen beskrivs av Bohlin ha varit bidragande till förändringen av RCTens status. McMastergruppen bestod av lärare som var verksam inom läkarutbildningen vid McMaster universitet i Toronto. Deras fokus var på att resultat från vetenskapliga studier skulle överordnas klinisk erfarenhet och expertbedömningar. Resultat från randomiserade kliniska prövningar ska enligt McMastergruppen premieras framför annan typ av vetenskapliga studier. Lärarna i McMastergruppen gick sedan samman med kollegor från andra lärosäten och bildade EBMWG. EBMWG (1992, ss. 2420-2421) skriver i början av 1990-talet att ett nytt paradigm för medicinsk praxis håller på att växa fram – evidensbaserad medicin. EBMWG lyfter fram att det är mycket mer troligt att randomiserade undersökningar leder oss rätt gällande frågor om ett läkemedel gör mer nytta än skada.

It is when asking questions about therapy that we should try to avoid the non-experimental approaches, since these routinely lead to false positive conclusions about efficacy.

Ovanstående citat synliggör att metoden är av betydelse för resultatets trovärdighet, men som vi kommer se längre fram innebär inte ett trovärdigt resultat att det per automatik anses vara överförbart till andra sammanhang.

### **Klinisk variation sätter strålkastarljuset på registerstudier**

Florence Nightingale beskrivs av Salive, Mayfield och Weissman (1990, s.699) som en av de första att bedriva outcomes research. I den kliniska praktiken förde hon noggranna register över sina patienters vård och utfallet av vården. Drygt ett halvt sekel efter Nightingales död, bidrog John Wennbergs forskning till att sätta strålkastarljuset på registerstudiers potential. Dock rådde det helt andra förutsättningar för att bedriva registerstudier för Wennberg jämfört med Nightingale. Vid mitten av 70-talet hade datorer kapacitet att lagra stora mängder av data, vilket möjliggjorde utveckling av stora databanker (Fries 1976, ss. 1400-1402). Wennberg använde sig av data från såväl försäkringsdokument som från sjukhusens utskrivningsdokument (1984, s. 15). Utifrån denna data påvisar han att det rådde stor variation vid såväl val av åtgärd som vid beslut om det behövs inläggande vård (Salive, Mayfield & Weissman 1990, s. 702; Wennberg 1984, s. 7). Wennberg (ss. 7-8) kommer fram till att variationen främst förekom för åtgärder kopplade till tillstånd som hör ihop med åldrandet. Den kliniska variationen beror enligt Wennberg på att de enskilda läkarnas egna subjektiva uppfattningar om vad som ska göras varierar, vilket han benämner "practice style factor". Wennberg rekommenderar därför regelbunden användning av register och databaser för att identifiera variationer inom de kliniska verksamheterna.

Wennberg använde inte bara registerstudier för att identifiera variation i val av åtgärd. Han hämtade data från Medicare (en försäkringsdatabas) för att studera utfallet av prostataoperationer i Maine, USA. Detta tillvägagångssätt, skriver Wennberg (s. 21, 24), ger läkarna i Maine den kunskap som är mest relevant för dem, nämligen hur utfallet för prostataoperationer i Maine ser ut. Han finner att mortaliteten post-operativt är mycket högre än vad som presenterats i den tidigare publicerade litteraturen. De flesta av de tidigare studierna hade studerat mortalitet under tiden som patienterna var

inneliggande, detta till skillnad mot Wennberg. Han hade följt patienterna i 2 år efter operationen. Wennberg (s. 24) beskriver att registerstudien inte fick stå oemotsagd och exemplifierar med det redaktionella svaret på artikeln. I det redaktionella svaret stod det att fyndet var atypiskt och kunde förklaras genom patientselektion och den metod som använts vid prostataoperationerna.

Precis som Bradford-Hill lyfter upp begränsade ekonomiska resurser som en förklaring till genomförandet av RCTen, lyfter Epstein (1990, s. 266-267) fram ekonomiska aspekter som bidragande faktor till outcomes movement (resultatrörelsens) framväxt. Detta kan förstås utifrån Relmans (1988, ss. 1220-1222) beskrivning av ökade kostnader kopplade till den amerikanska hälso- och sjukvården samtidigt som forskare som Wennberg synliggjorde att alla åtgärder som utfördes inom hälso- och sjukvården kanske inte var nödvändiga. Relman skriver 1988 att det finns ett behov av att veta mer om den variation som Wennberg synliggjort. Detta för att kunna sänka kostnaderna inom vården utan att samtidigt minska på dess kvalitet.

I likhet med studien om streptomycin kan användningen av registerstudier också ses stamma från ett behov av ökad kunskap om behandlingar och åtgärder, men här framkommer en betoning på effectiveness, alltså vilka effekter som återfinns i den kliniska verkligheten. Genom att registerstudier baseras på data som är hämtade från den kliniska praxisen, ses de som en spegling av den kliniska verkligheten. Resultat från registerstudier ses därför som relevanta då det ökar kunskapen om utfallen i den kliniska verkligheten, men precis som framkom i det redaktionella svaret på Wennbergs studie, finns det en osäkerhet vad resultatet i registerstudier egentligen står för. Det är något som vi kommer se längre fram i uppsatsen leder till att resultatets relevans för kliniska beslut ifrågasätts.

Epstein (1990) och Gerszten (1998) lyfter fram en aspekt som gör att outcomes research kan kopplas samman med ökad relevans, nämligen att utfallen i outcomes research hade ett bredare fokus. Det handlade inte längre bara om mortalitet, återinläggningar och komplikationer utan även om funktionell status, emotionell hälsa och patientnöjdhet. Att patientnöjdhet blev ett viktigt utfall kan åtminstone delvis förklaras av att det blev viktigt både vid val av försäkring som vid val av läkare och behandlingsalternativ, särskilt i en amerikansk kontext.

### **Valet mellan metodens stringens och resultatets relevans**

Fries (1976, ss. 1400-1402) lyfter upp problematiken med att kliniska frågor som man har för stunden vanligtvis aldrig har studerats systematiskt. Han skriver att antalet viktiga frågor är många, medan antalet besvarade frågor är relativt litet. Detta kopplar Fries samman med att de flesta prospektiva studier, som är utmärkande för RCTer, begränsas på olika sätt. Antalet individer som studeras är för få, uppföljningstiden är för kort, gruppen som studeras är högselektad och försökspersonernas följsamhet är orealistisk i en klinisk praxis. Fries ser minst tre sätt som databanker kan hjälpa kliniker i deras arbete. Databanker kan användas för att identifiera mindre produktiva aktiviteter, täcka upp för de saknade kliniska studierna samt ge beslutsstöd för hur de enskilda individerna ska handläggas. Databankernas innehåll bygger nämligen på klinisk erfarenhet. Att databanker kan hjälpa till att besvara kliniska frågor som inte studerats tidigare lyfter även Rosati, Wallace och Stead (1973, ss. 285-287) fram. De är också inne på samma linje som Fries, nämligen att databanker kan bidra till att en åtgärd anpassas till individen. Rosati et al skriver att patienter med en diagnos inte är en homogen grupp, utan snarare består av flera subgrupper med varierande prognos. De menar istället att en intervention för individer med samma diagnos kan leda till såväl bra, som neutral eller till och med dåligt resultat beroende på vilken subgrupp individen tillhör. Databanker medför en möjlighet att få tillräckligt stora populationer för att på ett meningsfullt sätt kunna studera de olika subgrupperna. Byar (1980, s. 342) lyfter dock ett varnande finger då han i sin artikel skriver:

It has been argued that data banks can address a wider range of questions than randomized clinical trials can. The big question, however, is whether data banks can answer them.

Byar (ss. 337-338) exemplifierar detta genom sin egen erfarenhet av ett register för sköldkörtelcancer. Det stora bekymret som Byar lyfter upp är att även om skillnaderna vid de flesta

behandlingsjämförelser är medicinskt viktiga, är de relativt små jämfört med den möjliga närvaron av systematisk bias. När en behandlings effekter studeras genom en registerstudie råder det ofta en betydande osäkerhet sett till om förbättring beror på behandlingen eller om resultatet istället beror på att de patienter som fått behandlingen redan innan hade en bättre prognos. Det finns alltså en osäkerhet kring vad resultaten från registerstudier egentligen står för. Detta till skillnad från RCTer som just undviker selektionsbias genom randomisering.

Förutom selektionsbias lyfter Byar (ss. 339-340) fram flera andra aspekter som gör det svårt att kunna veta vilka konklusioner som ska dras utifrån registerstudiens resultat. Det kan vara svårt att veta exakt vad som jämförs med vad. Exempelvis när kirurgiska åtgärder jämförs med medicinska behandlingar kan bäge grupperna innehålla en stor variation. Ett annat dilemma kan dyka upp vid studier där åtgärder från olika tidsperioder jämförs. Ifall diagnostiken förfinas kan det leda till att man i senare delen av registerstudien hittar cancer i ett tidigare skede än förr. Det kan då innebära att cancer som behandlas inte är lika allvarlig i den senare delen av studien, även om samma namn används. Detta kan snedvridera resultatet till fördel för den nyare behandlingsmetoden.

Att använda registerstudier för att se resultatet för subgrupper lyfter Byar (s. 341) upp som riskfyllt. Man riskerar nämligen då att hitta en effekt som egentligen inte finns utan som beror på slumpen. Byar ser istället andra användningsområden för databanker, t.ex. för deskriptiva studier, identifikation av prognostiska faktorer eller vid planering av RCTer. Hlatky (1991) å andra sidan har positiva erfarenheter av databanker. Hlatky (ss. 648-649) beskriver hur forskare har utvecklat och validerat en modell utifrån Duke Cardiovascular Disease Database. Modellen användes för att jämföra medicinsk behandling med kirurgiska åtgärder vid kranskärlssjuka. Förutsägelserna från modellen stämde väl överens med resultat från tre stora RCTer. Hlatky skriver att detta indikerar att när data är trovärdig och det finns en god förståelse för prognostiska faktorer, så kan forskare använda statistiska tekniker för att minimera selektionsbiasen. Hlatky skriver att modellen kan utgöra ett verktyg för att jämföra medicinska åtgärder med kirurgiska åtgärder för patienter som är underrepresenterade i RCTer såsom kvinnor och personer över 65 år.

I de genomgångna artiklarna framkommer spänningen mellan stringens och relevans. RCTerna skapar genom såväl inklusions- och exklusionskriterier samt genom randomisering två homogena subpopulationer. Det bidrar till att små men kliniskt viktiga skillnader mellan behandlingar kan synliggöras på ett trovärdigt sätt, samtidigt skiljer sig patienterna i RCTerna sig åt från de som man möter i kliniken. I RCTerna har patienterna en högre följsamhet, är yngre och är oftare man, detta till skillnad från patienterna i registerstudierna. De kan istället ses som en spegling av de patienter som man möter i det vardagliga arbetet, vilket kan förstås utifrån att registerstudier har en större spännvidd sett till vilka som inkluderas. I registerstudierna inkluderas både fler kvinnor och studierna består har en större andel äldre patienter. Registerstudiens population är inte så homogen som RCTens. I registerstudierna återfinns även en osäkerhet kopplat till vem och vad som studeras, samtidigt finns det faktorer som kan öka registerstudiernas trovärdighet. Bland dessa faktorer återfinns tillförlitligheten av datan som registerstudien grundas på samt vilken kännedom som finns sedan tidigare om tillståndet eller sjukdomen som studeras. Om detta sedan räcker för att resultatets trovärdighet ska komma i nivå med resultat från RCTer råder det delade meningar om.

## 5.2 Debatten fortsätter under det nya millenniet

Även efter millennieskiftet återfinns olika sätt att se på RCTer och olika typer av observationsstudier. Detta tycks stundom kunna väcka känslor, vilket kan anas i en av kommentaren till Andréasson et al artikel (2016). I sin artikel ifrågasätter Andréasson et al bevisvärdet av de observationsstudier som lyft fram att måttlig alkoholkonsumtion minskar risken för hjärtsjukdom. Fagrell (2016) kommenterar detta med:

*Återigen kommer det en artikel från Sven Andréasson och medarbetare som totalt nedvärderar alla de hundratals 'observationsstudier' som visat att en lätt till måttlig alkoholkonsumtion hos*



*medelålders och äldre påtagligt minskar risken för t ex hjärt-kärlsjukdom, diabetes typ2 och reumatoid artrit. Att alla dessa studier anses vara värdelösa är minst sagt provocerande!*

## **Randomiseringens styrkor och svagheter**

Andreasson et al (2016) ifrågasätter observationsstudiernas bevisvärde utifrån att studierna har metodologiska begränsningar. En av de viktigaste aspekterna är att exponeringen inte är randomiserad. I likhet med Andreasson et al lyfter Taube (2008) fram att RCTer har den stora fördelen att de är just randomiserade. Taube (s. 535) skriver att registerstudier kan ge en viktig indikation om utfallet i den kliniska praktiken, men det går inte att säga vad som är orsak till resultatet. Taube (ss. 534-537) exemplifierar det genom en fiktiv jämförelse mellan två olika kliniker som bägge utför en viss kirurgisk åtgärd. Finner man genom en registerstudie att resultat mellan klinikerna skiljer sig statistiskt åt, så går det inte att direkt dra slutsatsen att det finns en verklig systematisk skillnad i operationsskicklighet. Det kan istället vara så att den ena kliniken är skickligare på att välja ut patienter som är lämpliga för operation. En annan möjlighet är att patienterna skiljer sig åt vid de respektive klinikerna sett till olika prognostiska faktorer. I så fall kan inte grupperna ses som jämförbara. Taube drar därför slutsatsen att registerstudier inte kan utgöra bevis. Dessa studier är snarare hypotesgenererande istället för hypotesprövande, detta till skillnad från RCTer som ger resultat som är entydigt tolkbara. Individerna som opererats på de bägge klinikerna är i RCTen allokerade enligt slumpens lagar, vilket leder till att resultatet inte kan förklaras av skillnader mellan de två grupperna. Taube (s. 535) skriver:

*Detta innebär att inga systematiska skillnader mellan dessa grupper kan förväntas snedvrida analysresultatet – grupperna är jämförbara.*

Grossman och Mackenzie (2005, ss. 525-528) skriver att antagandet att randomisering alltid tar hand om bias skapar en illusion av homogenitet. Detta kan i sin tur leda till att man missar potentiella felkällor. Randomisering beskrivs som meningslöst när provstorleken är liten sett till antalet variabler. Detta då det är irrationellt att tro att små studier kan få välbalanserade grupper. Grossman och Mackenzie exemplifierar det genom att beskriva en studie där fyra individer, varav två är kvinnor, ska randomiseras. Om individerna slumpas till två grupper finns det en betydande risk att kvinnorna och männen kommer hamna i varsin grupp. Randomisering innebär alltså inte per automatik att grupperna blir välbalanserade. Grossman och Mackenzie lyfter även upp att det är viktigt att veta vilken enhet som ska randomiseras. Om man ska genomföra en intervention inom skolan är det viktigt att veta om randomisering ska ske på individnivå, klassnivå eller skolornivå. Sker randomisering på fel nivå kan poängen med randomisering försvinna, dessutom påverkar nivån hur stor provstorleken är. Är det individer som ska randomiseras kan det röra sig om några tusen medan är det skolor som ska randomiseras kanske det istället rör sig om tre till fyra enheter.

Grossman och Mackenzie (2005, ss. 518, 529) skriver att RCT är den bästa studiedesignen i vissa fall. Detta gäller exempelvis för att studera de primära behandlingseffekterna i stora fas III läkemedelsstudier. I andra fall när det t.ex. gäller folkhälsointerventioner är observationsstudier inte bara mer görliga, utan ger även mer exakta resultat. Författarna ger även exempel på medicinska frågor där randomisering inte är att föredra. Detta gäller exempelvis botulinumtoxin som behandling vid muskelspasticitet efter stroke. RCTen ses då som en mindre passande studiedesign då denna studiedesign har sämre förmåga att undersöka *“the impact of the intervention on functional abilities and social participation”* (s. 529). Grossman och Mackenzie lyfter även fram att det inte alltid är etiskt möjligt att randomisera, exempelvis när man studerar effekter av rökning. Även Rosén, Axelsson och Lindblom (2008, s. 3191) ser att det kan finnas etiska bekymmer med randomisering. Pap-smear-screening startade utan en RCT, men där minskningen i incidens av livmoderhalscancer på befolkningsnivå har varit så kraftfull att effekterna inte kan förnekas. Rosén et al ser det därför oetiskt att nu starta en RCT.

Howick (2011, ss. 36, 50-51) lyfter upp att oavsett studiedesign kan två grupper aldrig vara exakt lika, detta då det finns oändligt många skillnader mellan grupperna. RCTer kan alltså inte utesluta alla

confounders, men RCTer utesluter fler confounders än observationsstudier. Randomisering utesluter selektionsbias, vilket inte kohortstudier gör. Randomisering medför dessutom möjlighet för dubbelmaskerade, placebo-kontrollerade studier och har därmed även potential att minska behandlingsbias. Howick kommer därför fram till att i enlighet med regeln att evidensens kvalitet beror på hur väl confounders har uteslutits, ifall alla andra saker är lika, så är evidens från RCTer överlägsen evidens från kohortstudier.

Beskrivningarna av randomisering lyfter upp både dess styrkor och dess svagheter. Randomisering beskrivs bidra till att resultatet har ökad trovärdighet, vilket medför att studier som inte randomiserar deltagarna kan ifrågasättas sett till vad resultatet egentligen står för. Randomisering behöver dock inte per automatik leda till att den interna validiteten ökar, exempelvis kan fel enhet randomiseras. Det är inte heller säkert att den förväntade homogeniteten mellan olika grupper skapas genom randomisering. Det kan även vara så att viljan och behovet att randomisera studiedeltagare begränsas av såväl etiska principer som av vilka svar man söker.

### **Effektstorlek: observationsstudiers möjlighet till övertag**

Howick (2011, ss. 39-40) skriver att EBM-hierarkien där RCTer är placerade över observationsstudier skapar en paradox. Behandlingar som ses som de mest effektiva, t.ex. Heimlichmanövern vid blockerad luftväg, har aldrig testats i RCTer. Howick konstaterar därmed att effektiva behandlingar inte stöds av den bästa evidensen. För att lösa denna paradox lyfter Howick (ss. 56-61) fram effektstorleken (*effect size*) som en viktig faktor för att bedöma evidensstyrkan. Begreppet effektstorlek använder Howick för att benämna den observerade effekten av en åtgärd i en jämförande studie. Howick föreslår att evidenshierarkin ska ersättas med en bedömning av effektstorleken, vilket då innebär att observationsstudier ger god evidens när effektstorleken är större än effekten av möjliga confounders. Liknande tankegångar om att effektstorleken är av betydelse återfinns i GRADE. GRADE är ett internationellt system som används för att gradera evidensstyrkan, vilket alltså är en gradering av styrkan på det vetenskapliga underlaget. Evidensgraderingen i GRADE består av fyra steg, från den lägsta graderingen vilket benämns mycket lågt vetenskapligt underlag till den högsta graderingen vilken benämns starkt vetenskapligt underlag. Evidensgradering görs separat för varje enskilt effektmått. En sammanvägning enligt GRADE görs efter att de enskilda studierna kvalitetsgranskats. Först görs en preliminär bedömning av evidensstyrkan, vilket enbart grundas på studiedesign hos de studier som granskas. RCTer innebär att den preliminära styrkan ses som stark (högsta graderingen) medan för observationsstudier är den preliminära evidensstyrkan låg (näst lägsta graderingen). Efter den preliminära bedömningen görs en bedömning av olika faktorer som både kan medföra uppgradering liksom nedgradering av evidensstyrkan. Faktorer som bedöms är bl.a. överförbarhet, effektstyrkan, precision och sannolikhet att effekten är underskattad. Det innebär att observationsstudier ges möjlighet att uppgraderas om effektstyrkan är tillräckligt stor (relativa risken (RR) är  $>2$  eller  $<0.5$ ) (SBU 2012; SBU 2014b, ss. 44, 141-144, 148). Howick skriver att fokus på effektstorlek möjliggör att en välgjord, kontrollerad observationsstudie som visar på en stor effektskillnad kan ge starkare evidens än en RCT med små effekter. Howick (2011, ss. 56-59) beskriver dock att han går steget längre än GRADE, då han anser att RCTer ska kunna nedgraderas om studierna visar på väldigt små effekter. För att stödja sitt resonemang lyfter han upp studier av SSRI-preparat (medicin mot nedstämdhet och ångest). Den absoluta effekten för SSRI-preparat i jämförelse med placebo beskriver han som ringa 6 % (liten effektstorlek), samtidigt som det finns en bortglömde källa för confounding. Läkemedlets biverkningar möjliggör för patienterna att veta om de fått den aktiva behandlingen eller placebo. Detta utgör därför enligt Howick ett exempel på när åtgärdens effektstorleken inte behöver vara större än den kombinerade effekten av möjliga confounders. Det föreligger bias som skulle kunna förklara studiens resultat.

Concato, Shah och Horwitz (2000) har även de intresserat sig för effekter. Deras utgångspunkt är att undersöka det som de beskriver som en allmänt accepterad sanning, nämligen att observationsstudier i allmänhet övervärderar effekten av interventioner jämfört med RCTer. För att belysa observationsstudiers plats i utvärdering av behandling genomför de en jämförelse mellan RCTer och

observationsstudier med samtida kontroller. De analyserar fem kliniska områden utifrån 99 rapporter. Det genomsnittliga resultatet för observationsstudierna beskrivs vara "*remarkably similar to those of the randomized, controlled trials*" (s. 1887). De studerar bl.a. BCG-vaccination för att förhindra aktiv tuberkulos. De tretton inkluderade RCTerna har en relativ risk på 0.49 (95%KI 0.34-0.70) att få aktiv tuberkulos hos de som blivit vaccinerade jämfört med de som inte vaccinerats. Detta jämförs med resultatet från tio fall-kontrollstudier där oddskvoten är 0.50 (95%KI 0.39-0.65). Fall-kontrollstudierna visar alltså på liknande riskreducerande effekt vid BCG-vaccination som RCTerna. Förutom att resultaten är liknande, lyfter Concato et al (s. 1888) fram att punktestimaten för de enskilda studierna har större spridning bland RCTerna (0.20-1.56) jämfört med observationsstudierna (0.17-0.84). Det innebär att de enskilda resultaten i de olika RCTerna skiljer sig mer åt än vad det gör för de olika observationsstudierna. Att observationsstudier har mindre benägenhet för heterogenitet förklarar författarna (s. 1890) genom att dessa studier inkluderar en bred population. RCTer får istället mer distinkta studiepopulationer p.g.a. deras inklusions- och exklusionskriterier, vilket leder till att resultaten i de olika RCTerna skiljer sig åt mer än observationsstudierna.

Concato et al (s. 1890-1892) drar slutsatsen att väldegnade observationsstudier såväl kohortstudier som fall-kontrollstudier inte systematiskt överskattar effektstorleken av behandling jämfört med RCTer. De lyfter fram att det är viktigt att studierna har liknande exklusionskriterier och att prognostiska faktorer tas med i beräkningen när resultat från observationsstudier ska jämföras med RCTer. Concato et al skriver att den vitt spridda föreställningen att enbart RCTer kan producera trovärdiga resultat inte är till gagn varken för patienter eller personal.

Pocock och Elbourne (2000) ställer sig dock inte bakom denna slutsats. De ser istället en fara för patienters välbefinnande ifall observationsstudier skulle öka i användning när behandlingar utvärderas (s. 1907). Pocock och Elbourne skriver att det är svårt att bedöma trovärdigheten av Concato et als studie. Detta då det är svårt att utifrån Concato et als artikel granska de enskilda studierna som deras artikel bygger på (s. 1908). Pocock och Elbourne har även andra invändningar till studiens utformning. De ifrågasätter bl.a. valet av meta-analys när screening av bröstcancer undersöktes. I den inkluderade meta-analysen hade RCTernas kvalitet inte haft betydelse. Concato et al (s. 1891) skriver dock i sin artikel att när de kontrollerade för studiernas kvalitet förändrades inte resultatet. Pocock och Elbourne (s. 1908) lyfter även upp en nyare meta-analys som inte visar på samma effekter gällande bröstcancerscreening som de inkluderade meta-analyserna i Concato et als artikel gör. Pocock och Elbourne (s. 1908) drar slutsatsen att det är troligt att de inkluderade rapporterna är högt selekterade eftersom det är ovanligt för en terapeutisk fråga att vara undersökt både genom observationsstudier och genom experimentella studier. Enligt Pocock och Elbourne (s. 1907) så är det endast genom randomisering som trovärdiga estimat av behandlingseffekt kan fås. Även om man kan justera för skillnader mellan grupper i observationsstudier, så är det omöjligt att veta vilka som är det adekvata justeringarna eller om man har dokumenterat all relevant data om patienterna.

Effektstorleken framträder som en faktor som kan bidra till att resultatets trovärdighet såväl minskar som ökar. Det är även intressant att se hur RCTens homogena studiepopulation som tidigare beskrivits som ett sätt att öka resultatets trovärdighet, samtidigt kan leda till en ökad heterogenitet när resultat mellan olika RCTer jämförs. Homogenitet som tidigare framställts som ett signum för att beskriva RCTens studiepopulation, används nu när olika registerstudiers populationer jämförs med varandra.

### **Kvalitetsgranskningens betydelse vid bedömning av studiernas trovärdighet**

Rosén, Axelsson och Lindblom (2008) använder Concato et als (2000) artikel som ett stöd för att visa att observationsstudier inte per automatik ska anses mindre tillförlitliga än RCTer. Rosén et al (2008, s. 3192) utgångspunkt är att kvalitetsgranskning och bedömning av tillförlitlighet är viktigt för såväl RCTer som observationsstudier. Problem som de lyfter upp gällande RCTer är snäva inklusionskriterier, bristande randomisering och att ekonomiska intressen riskerar att leda till systematiska fel när studien finansieras av industrin. För observationsstudier skriver de att man ska

vara försiktig i sin tolkning av resultatet när studierna är små, baserade på historiska kontroller, där utfallsmåttet är självrapporterat och där kontroll för förväxlingsfaktorer saknats eller varit bristfälligt.

Vikten av att ha kännedom om potentiella förväxlingsfaktorer och ta med dem i beräkningen framkommer i observationsstudier som studerat ett högt intag av vitamin E och C respektive östrogenbehandling för kvinnor i klimakteriet. Bägge dessa åtgärder bedömdes utifrån observationsstudier kunna leda till minskad risk för hjärt-kärlsjukdom (Rosén et al, s. 3192). När dessa åtgärder senare undersöktes genom randomiserade kontrollerade undersökningar visade det sig att dessa åtgärder inte var kopplade till minskad risk för hjärt-kärlsjukdom. Rosén et al konkluderar att orsaken till att observationsstudier hade påtalat att en åtgärd var effektiv när den inte var det, berodde på att dessa studier inte tillräckligt noga hade kontrollerat för förväxlingsfaktorer. Riskminskningen i observationsstudierna försvann för såväl vitaminerna som för östrogen när hänsyn i efterhand togs till socioekonomiska faktorer. Rosén et al lyfter fram att (s. 3192):

Sociala och ekonomiska faktorer är sannolikt förutom ålder de mest betydelsefulla riskfaktorerna för förtida död och sjuklighet när det gäller många sjukdomar.

De skriver att det därför är särskilt viktigt att kontrollera för socioekonomiska faktorer när man studerar preventiva, symtomlindrande och icke-akuta insatser. "*Välutbildade personer med god social position*" har mer kunskap om vilka möjligheter som står till buds och de har även en ökad förmåga att få igenom sina krav (Rosén et al s. 3193). Det är inte bara observationsstudier som riskerar att leda i fel riktning. Rosén et al beskriver hur läkemedlet aprotinin användes under 14 år för att minska blodförluster vid bypass-kirurgi. Trots att två observationsstudier visade på en högre dödlighetsrisk för patienterna skickade FDA (Food and Drug Administration, läkemedelsmyndighet i USA) inte ut någon varningssignal, eftersom RCTer inte synliggjort denna risk. Det var först när en stor RCT fick stoppas i förtid som läkemedlet drogs in. Rosén et al ser ändå att välgjorda och stora RCTer ger de säkraste resultaten när det gäller att uppskatta behandlingseffekter.

Rosén et al synliggör behovet av att bedöma studiernas kvalitet, för att på det sättet kunna bedöma resultatets trovärdighet. För registerstudier ter det sig som extra viktigt att justera för socioekonomiska faktorer, åtminstone om studien har som ambition att inte bara beskriva den kliniska verkligheten utan även syftar till att trovärdigt uttala sig om samband mellan åtgärd och effekt. Det är samtidigt lika viktigt att RCTer granskas för att se om de på ett trovärdigt sätt kan besvara den fråga man söker svar på.

## Relevans genom överförbarhet

Jernberg och Janzon (2015) skriver att randomiserade kliniska prövningar är det bästa sättet att jämföra två behandlingsalternativ, samtidigt som de är av åsikten att registerdata bör användas i större utsträckning vid utvärdering av behandlingar. Bekymret med RCTer är enligt dem att patienter selekteras både genom studiernas inklusions- respektive exklusionskriterier, men också då enheter som rekryterar patienter har ytterligare faktorer som påverkar vilka som inkluderas i studien. Det kan finnas konkurrerande studier samt egna preferenser hos de som rekryterar deltagare. Detta leder fram till att RCTer skiljer sig åt från den kliniska praxisen. Patienter i den kliniska verkligheten är ofta äldre (det kan röra sig om en skillnad på 10 år enligt Jernberg och Janzon). Denna åldersskillnad leder till att de är färre kvinnor, lägre andel med starkt ogynnsamma prognostiska faktorer (ex. tidigare stroke, hjärtsvikt och njursvikt) i de randomiserade studierna. Det kan vara problematiskt när man ska överföra resultaten från studierna till den kliniska praxisen. Jernberg och Janzon skriver att i kliniska prövningar där högrisk patienter är inkluderade var andelen med kardiovaskulär död, hjärtinfarkt eller stroke under de följande åren ungefär hälften av det vi ser i den kliniska verkligheten. Den kliniska verkligheten utgörs i deras exempel av data från en registerstudie som följt upp utfallet efter hjärtinfarkt. Jernberg och Janzon skriver att (2015, s. 1):

I verkligheten kan både de positiva och de negativa effekterna vara större i absoluta mått mätt, och balansen mellan nytta och skada kan se annorlunda ut.

Norlin, Schmitt-Egenolf, Steen Carlson och Persson (2014) lyfter upp att det kan vara skillnad mellan den kliniska praxisen och RCTer. Skillnaden beskrivs kunna bero på flera olika faktorer. Precis som Jernberg och Janzon (2015) lyfter de upp att det kan föreligga skillnad i patientklientelet. Individerna i den kliniska verkligheten är äldre, har oftare samsjuklighet eller ovanliga symtom till skillnad från individer som är med i RCTer. De lyfter dessutom fram att dosen av läkemedlet kan ha justerats av läkaren för att passa individen. Det kan i den kliniska praxisen även föreligga en indikationsglidning då läkemedlet förskrivs till en annan patientgrupp än de som var avsedda i indikationen. De exemplifierar detta med fynd från granskning av psoriasisregistret. Det var inte enbart de med måttlig till svår psoriasis som fick nya biologiska behandlingar utan även patienter med mildare psoriasis. Indikationsglidning kan leda till att den effekt man förväntar sig av läkemedlet inte blir så stor, då det är en patientgrupp med mildare symtom som behandlas. Den kliniska verkligheten är alltså inte så strikt kontrollerad som RCTer är. Norlin et al (2014) anser därför att nationella kvalitetsregister är viktiga för att ge information om nya läkemedels effektivitet d.v.s. hur de fungerar i klinisk vardagen. Registrena kan även bidra med information om hur läkemedel används i klinisk praxis.

Howick (s. 54-55) skriver att observationsstudier liksom de flesta randomiserade undersökningarna ger ett medelvärdesresultat. Det går inte att rakt av appliceras på den enskilda individen, ifall det inte förekom en liten variation i behandlingssvaret. Howick skriver att observationsstudier generellt sett inte är mer representativa än RCTer. Även om det i många fall är så att RCT exkluderar deltagare som sedan erhåller terapin i klinisk praxis, är det inte ett nödvändigt drag hos de randomiserade undersökningarna. Taube (2008) ser däremot inklusions- och exklusionskriterierna som ett verktyg som bidrar till att RCTer producerar resultat som är trovärdiga. Enligt Howick (s. 54-55) så är frågan om huruvida en studie har extern validitet inte av intresse ifall studien inte har god intern validitet. Att överföra resultat från en studie till en annan population är meningslöst så länge som resultatet inte är trovärdigt.

Howick belyser att både registerstudier och RCTers studiedesign leder fram till medelvärdespatienter. Patienterna i registerstudier ses däremot ha ovanligare symtom, ha fått individanpassad dos av läkemedlet, är inte lika kontrollerade och kan ha fått ordinerat läkemedlet på annan indikation än som var tänkt från början. Intressant nog ses registerstudier inte bara spegla den kliniska verkligheten utan beskrivs i vissa artiklar vara den kliniska verkligheten. I RCTer beskrivs patienterna som friskare (inte lika vanligt med stroke, hjärtsvikt och njursvikt) än patienter i kliniska praxisen.

### **Går det att välja både stringens och relevans?**

RCTer och registerstudier har både vissa likheter men också skillnader sett till deras studiedesign, vilket kan leda till att en studieform premieras på bekostnad av den andra. Det finns samtidigt exempel på hur man försöker kombinera de bägge studieformernas styrkor. Här presenteras ett exempel på hur det kan göras i själva genomförandet av studien samt ett exempel på hur det kan göras när studierna är genomförda.

#### ***Registerbaserad RCT***

Jernberg och Janzon (2015) skriver att randomiserade studier måste göras närmre den kliniska verkligheten. De lyfter fram registerbaserade RCTer (RRCT) som en möjlig väg att vandra. Enligt Fröbert och Cizinsky (2014) är TASTE (Thrombus Aspiration during ST-Elevation Myocardial Infarction in Scandinavian) den första RRCT i världen. I denna RRCT av Fröbert et al (2013) undersöks om åtgärden att suga ut blodproppar från hjärtats kranskärl innan ballongvidgning (Percutan Coronary Intervention, PCI) medför fler räddade liv vid akut hjärtinfarkt. Fröbert et al (2013) beskriver att evidensbasen, innan deras studie genomfördes, var något tvetydig. Tidigare studier pekade både på överlevnadsvinster och ökad risk för stroke. Bekymret med de tidigare RCTerna beskrivs av Fröbert et al (2013, s. 1588) som att de "have not generally been powered for hard clinical end points". I TASTE använde man sig av ett nationellt register som plattform för randomisering, insamlande av sjukdomshistoria och uppföljningsdata. I studien ingick alla 29 PCI-center i Sverige, samt även ett PCI-center från Island respektive Danmark. Studien omfattade drygt 7200 patienter. I

studien presenteras både baslinjedata för de som randomiserades men också för de som exkluderades innan randomisering gjordes. Man presenterar även uppföljningsdata för alla patienter i Sverige (även för de som inte randomiserades). Det synliggjorde att de som exkluderades hade en högre mortalitet (låg på cirka 10 %) jämfört med de som randomiserades (låg på knappa 3 %). En förklaring som ges är att deltagarna var tvungna att kunna ge muntligt godkännande för att få delta i studien, vilket innebär att de som var medvetlösa inte kunde vara med. Detta ser Fröbert et al (2013) som en av studiens begränsningar. De lyfter fram att även en studie som använder sig av ett register, inte fullt ut kan vara representativt för den kompletta bredden av patienter. Ett dilemma som Fröbert och Cizinsky (2014) sedan skriver en fördjupad artikel om. De skriver att:

Dagens etiska forskningsregler är anpassade till konventionella randomiserade studier, men är inte optimala för att möta de etiska problem som uppstår vid registerrandomiserade studier.

Fröbert och Cizinsky lyfter upp att innan TASTE-undersökningen genomfördes så användes trombosaspiration vid cirka 50 % av alla ingrepp. Beslut togs då av den enskilda operatören, utifrån kunskap som baserats på små studier av varierande kvalitet. Det fanns alltså varken ett tydligt stöd för eller emot proceduren. Om trombosaspiration skulle göras eller inte, var inte heller något som patient tillfrågades om i den kliniska verksamheten.

När TASTE planerades fick man inte inkludera patienter som inte kunde lämna ett medgivande, exempelvis patienter med uttalad medvetandepåverkan. Fröbert och Cizinsky (2014) skriver:

Detta gäller huvudsakligen patienter med kardiogen chock, där en delvis annan patofysiologi skulle kunna innebära att de kanske får mest, eller minst, effekt av trombos aspiration som tillägsbehandling. Ansökan avslogs av både den regionala och den centrala etiska kommittén. Vår uppfattning är att avslaget att randomisera de svårast sjuka var oetiskt eftersom TASTE har medverkat till ändrade riktlinjer för infarktbehandling utan att kunna ge klara rekommendationer för behandlingen av patienter i kardiogen chock.

Fröbert och Cizinsky lyfter upp ett möjligt tillvägagångssätt i framtiden, där inspiration hämtas från en brittisk studie. Det innebär att informerat samtycke inhämtas efter randomisering och genomförd åtgärd. Fördelen med detta tillvägagångssätt förutom att de sjukaste patienterna då kan inkluderas är att patienten är under mindre stress vid beslutstillfället. Trots att de svårast sjuka inte kunde inkluderas ser ändå Jernberg och Janzon (2015) att RRCT inkluderar patienter mindre selektivt än RCTer. RRCT innebär dessutom en lägre kostnader och därmed behöver frågeställningarna inte styras i samma grad av läkemedelsföretag (som tar marknadsmässig hänsyn i design av studier).

I varsin powerpointpresentation av James och Jacobsson (u.å.) respektive Lagerqvist (u.å.) beskrivs att RRCT, precis som i TASTE-studien, kan användas för värdering av strategier och behandlingar som redan används i klinisk rutin. När det är oprövade läkemedel och tekniska hjälpmedel som ska studeras behövs däremot traditionella RCTer. Detta håller även Jernberg och Janzon (2015) med om då de skriver att det i en tidig utvecklingsfas finns ett behov av den extra noggranna uppföljning som finns i RCTer. Register kan i dessa fall istället användas för att kunna identifiera patienter att inkludera i RCTen. När det handlar om ett läkemedel som redan används i kliniken, men där det kan vara aktuellt för en ny indikation kan såväl RRCTer som RCTer vara aktuellt (James & Jacobsson; Lagerqvist).

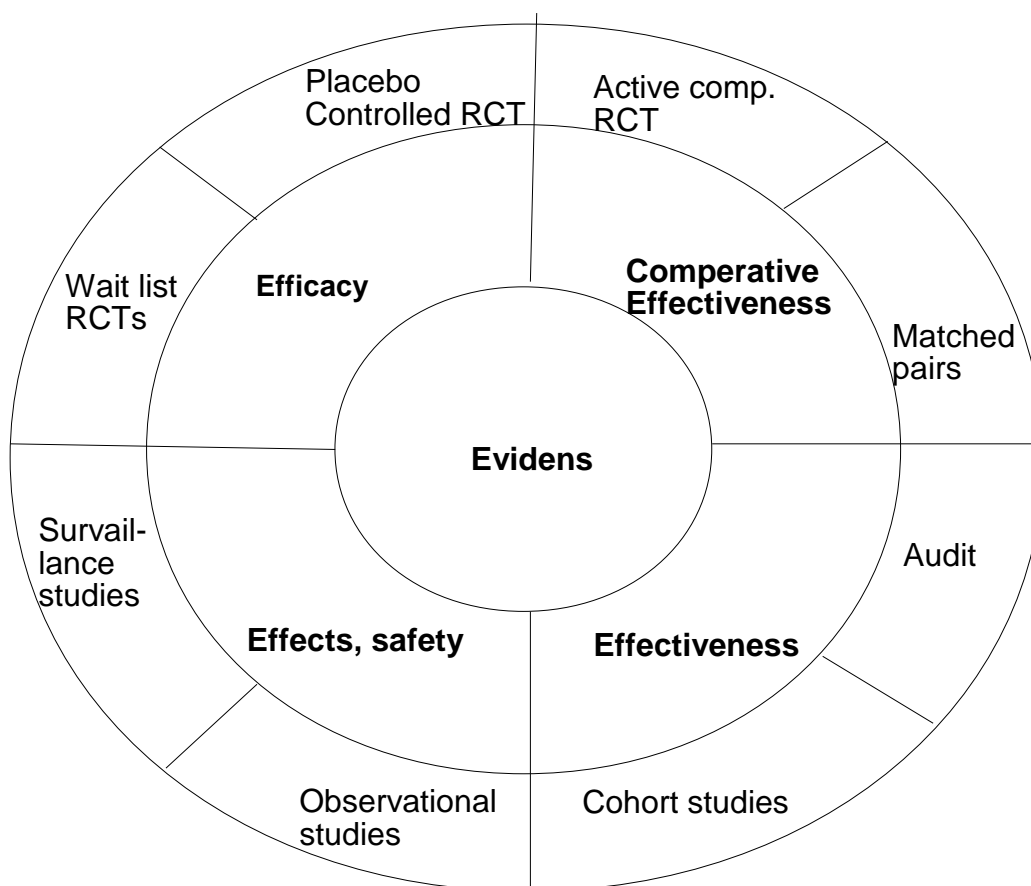
Mikael Åström, t.f. enhetschef Epidemiologi och registercentrum i Region Skåne, tog under läkemedelskongressen i Stockholm 2015 upp frågan om RRCT. Han betonade att det kan finnas en risk att man går för fort fram vid introduceringen av denna nya studieform. Han lyfte upp vikten av att registret som används innehåller kompletta och pålitliga variabler. Vid kliniska prövningar för man in data flera gånger och kontrollerar att det verkligen stämmer. För register kan det istället handla om en sekreterare som matar in data för hand och utan kontroller. Dessutom tog Åström upp att det är viktigt att det finns en enighet i registernätverket (Toresson 2015). Datans pålitlighet ser också Thörn och Järhult (2015) som ett dilemma. De skriver att det kan ske ett urval av vad och vilka som klinikerna rapporterar till kvalitetsregisterna, alltifrån vem som registreras till ren manipulation. Detta

både för att framstå i god dager och för att det kan finnas ekonomiska motiv. Enligt Thörn och Järhult är detta problem också något som har erkänts av myndigheter däribland riksrevisionen.

RRCT innebär alltså att stringensen uppehålls genom randomisering, medan kliniska frågor och studiepopulation som speglar den kliniska verkligheten bidrar med relevans. Det kan samtidigt anas orosmoln på RRCT-himlen i form av registerdatans tillförlitlighet och etiska dilemman.

### *Cirkulär evidensmodell*

Walach, Falkenberg, Fonnebo, Lewith och Jonas (2006, s. 5) skriver att varje metod har sina styrkor och svagheter, vilket inte kan lösas inom själva studiedesignen. Lösningen är istället att kombinera olika typer av studier när de väl är genomförda. De föreslår att evidenshierarkin ersätts med en cirkelmodell. Figur 1 visar cirkelmodellen (designad efter den modell som presenteras i Walach et al artikel från 2006, s. 5). Walach et al skriver att en cirkulär evidensmodell leder till en pragmatisk men fortsatt rigorös evidensbas. Deras utgångspunkt är att evidenshierarkin som används för att rangordna studier gällande läkemedel, inte fungerar när den överförs till andra områden såsom bedömningar av interventioner inom kirurgi eller psykoterapi. Evidenshierarkin bygger på ett visst antal antaganden, som sällan debatteras. (ss. 2-4). Dessa antagande gäller sällan för icke-medicinska interventioner. Även när läkemedel är i fokus är det sällan som alla antaganden är uppfyllda, så cirkelmodellen passar även för frågor gällande läkemedel. Bland dessa antaganden återfinns att patient och vårdgivare inte har några preferenser sett till val av behandling. Walach et al lyfter även upp antaganden om att det finns en sann, stabil effektstorlek som är oberoende av kontexten samt att de enda värdefulla effekterna är de som går att hänföra till en väl specificerad mekanism.



**Figur 1** Cirkelmodellen gjord efter Walach et al artikel (2006)

Genom triangulering, vilket innebär att olika typer av studiedesign används för att besvara en fråga, fås en bättre evidensbas, eftersom resultat som fås genom olika metoder jämförs med varandra.

Genom sin utformning synliggör cirkelmodellen att det inte finns en bästa metod. Utifrån cirkelmodellen blir inte den viktigaste frågan om studien är randomiserad. Fokus är istället på om man använt sig av en metod som är anpassad för frågan och om metoden har implementerats med optimal vetenskaplig stringens, då det handlar om att balansera olika studiers svagheter och styrkor. Cirkelmodellens övre halva innefattar experimentella studier vilka ofta ses ha högre intern validitet. I den övre halvan av modellen ingår placebokontrollerade RCTer och RCTer med aktiv kontrollgrupp. Metoderna i den nedre halvan kan bidra med extern validitet då de kan svara på frågor om säkerhet, användningsmönster och effectiveness, här återfinns t.ex. observationsstudier. Någon närmare beskrivning av audits och surveillance studies ges inte i deras artikel.

Sammanfattningsvis så leder cirkelmodellen till att fokus inte är på ifall studien är randomiserad eller inte. Fokus är istället på om metoderna används till rätt typ av frågor och om de genomförs på ett korrekt sätt. Olika kvantitativa metoder ses som jämbördiga, förutsatt att de tillämpas på rätt sätt.

Walach et al är inte ensamma om att se evidenshierarkin som problematisk. Howick (2011, s. 58) skriver att hierarkin uppmuntrar forskare till att förkasta observationsstudier när RCTer finns tillgängliga. Howick skriver att förutsatt att observationsstudierna uppfyller vissa specifika krav kopplat till effektstorlek, finns det ingen anledning till att inte väga samman resultat från observationsstudier med RCTer när man ska avgöra om en behandling ger kliniskt relevanta effekter.

### 5.3 Design av patient och resultat i enskilda studier

Granskningen av den enskilda studierna är av utforskande karaktär. De inkluderade artiklarna utgörs av fem registerstudier som alla använt sig av NDR (Nationella Diabetesregistret) samt av fem RCTer, vilka alla har kontrollgrupper. De inkluderade studierna har fokus på olika åtgärder, förebyggande såväl som behandlade, vid diabetes typ 2. I alla artiklar utom en ingår individer med diabetes typ 2. Undantaget är en RCT som studerar läkemedel för individer med pre-diabetes (förstadium till diabetes typ 2). Ytterligare en RCT och tre registerstudier fokuserar på olika läkemedel, men det är ingen studie som har fokus på de nya diabetesmedicinerna. En RCT respektive en registerstudier studerar utfallet efter operativa åtgärder. När det gäller livsstilsinterventioner fokuserar en registerstudie och en RCT på fysisk aktivitet samt en RCT som jämför två olika dieter. De inkluderade artiklarna presenteras i tabell 1.

Först kommer jag belysa design av patient genom en genomgång av hur patienter inkluderas och exkluderas samt vad det kan innebära för studien. Fokus är även på hur de inkluderade patienterna beskrivs. För att komplettera bilden har jag dessutom valt att se på design av resultat genom att se vad man studerar och vilka utfallsmått man valt. Jag kommer även presentera hur man tolkar utfallen, då det framkom en flexibilitet i detta mellan olika tillvägagångssätt.

#### Hur väljs patienter ut?

I flera av de granskade registerstudierna kan man se en ambition att vara rikstäckande. RCTer uppvisar en mindre urvalspopulation (den population som deltagarna rekryteras ifrån). För de studier som urvalspopulationen beskrivs, varierade den mellan en till två primärvårdsenheter upptill sex diabeteskliniker. Två av RCTerna inkluderade kliniker både inom och utom Sverige (Risstad et al 2015; Chiasson et al 2003). I Chiasson et als studie är nio olika länder med, men detaljer om hur rekrytering av deltagare skett framgår inte. Registerstudierna syftar ofta till att vara rikstäckande medan RCTernas urvalspopulation är mindre, vilket är en faktor som bidrar till studiepopulationernas storleksskillnad. I RCTerna varierar antalet studiedeltagare mellan 51-1429 medan motsvarande siffror för registerstudierna är 15 462 - 114 841.

Vid granskning av hur inklusions- respektive exklusionskriterierna används, framkommer en variation inom bägge typer av studiedesign. Det finns exempel på såväl snäva som vida kriterier. Exempel på en mer inkluderande RCT är Guldbrands et al (2012) studie där olika dieter jämförs. Deras inklusionskriterier var diagnosticerad med typ 2 diabetes med kostbehandling, oavsett om de hade



läkemedelsbehandling eller inte. De som exkluderades var patienter som hade svårt att förstå svenska språket, hade allvarlig psykisk sjukdom, malignitet eller missbrukade droger. Jonasson et als (2009) registerstudie som studerade association mellan malignitet och en insulintyp är ett annat exempel på en inkluderande studie. Individer som inkluderades var mellan 35-84 år och hade hämtat ut minst ett insulinrecept under inklusionsperioden. De som hade haft en malignitetsdiagnos innan studiestart uteslöts. Exempel på snävare kriterier återfinns i Risstad et als RCT (2015). Studien vände sig till

**Tabell 1** Presentation av granskade RCTer och registerstudier (förkortas i tabellen Reg.)

Förste författare (år)	Typ av studie	Studiens huvudsyfte	Studiens slutsatser	Antal deltagare	Citering /år
Risstad (2015)	RCT	Jämföra effekterna 5 år efter gastric bypass respektive duodenal switch*.	I jämförelse med gastric bypass resulterade duodenal switch i större viktminskning och större minskning av blodsocker. Duodenal switch var associerad med mer biverkningar.	61	10.0
Guldbrand (2012)	RCT	Jämföra effekterna efter 2 år med antingen low-fat diet (LFD) eller low-carbohydrate diet (LCD).	Vid jämförelse med den traditionella LFD är det ur hjärtkärl avseende säkert att sikta på 20% energi-intag från kolhydrater (= LCD).	61	6.8
Fritz (2011)	RCT	Bedöma stavgångs påverkan på livskvalitet hos överviktiga individer med typ 2 diabetes, normal eller nedsatt blodsockertolerans.	Stavgång kan introduceras i primärvården som en åtgärd som främjar viktminskning och livskvalitet till en låg kostnad.	212	5.0
Alvarsson (2003)	RCT	Undersöka huruvida behandling med insulin jämfört med glibenklamid (läkemedel) är associerat med bättre betacellsfunktion, hos individer som nyligen diagnostiserats med typ 2 diabetes.	I jämförelse med glibenklamid medför tidig insulin behandling vid diabetes typ 2, tillfällig förlängning av den endogena insulinsekretionen samt främjar bättre metabol kontroll.	51	8.2
Chiasson (2003)	RCT	Utvärdera effekten av att sänka blodsockret efter måltid med akarbos (läkemedel) hos patienter med nedsatt blodsockertolerans.	Behandling med akarbos hos patienter med nedsatt blodsockertolerans är associerat med signifikant riskreduktion för hjärtkärlsjukdom och hypertoni.	1429	60.9
Holzmann (2015)	Reg.	Undersöka långtidsöverlevnad efter CABG-operation** hos patienter med typ 1 diabetes respektive typ 2 diabetes.	Patienter med typ 1 diabetes hade mer än dubblad risk för att dö efter CABG-operation jämfört med patienter utan diabetes. Risken vara bara lätt ökad för patienter med typ 2 diabetes.	39 235	4.0
Zethelius (2014)	Reg.	Estimera risken av hjärtkärlsjukdom och dödlighet vid låg respektive hög nivå av fysisk aktivitet.	Studien visade på ökad risk för hjärtkärlsjukdom och mortalitet vid låg nivå av fysisk aktivitet.	15 462	7.5
Ekström (2012)	Reg.	Utvärdera effectiveness och säkerhet vid användning av metformin (läkemedel) i klinisk praxis.	Metformin visade på lägre risk jämfört med insulin för hjärtkärlsjukdom och totaldödlighet. Vid nedsatt njurfunktion sågs inga ökade risker.	51 675	9.0
Eliasson (2011)	Reg.	Beskriva användningen och utvärdera effectiveness av olika blodfettssänkande läkemedel hos patienter med typ 1 diabetes eller typ 2 diabetes.	LDL-C-nivåer (kolesterol) för patienter som använder simvastatin, atorvastatin eller rosuvastatin är väldigt lika liksom läkemedlens LDL-C sänkande effekt. Finns potential att intensifiera kolesterolsänkande behandlingen.	37 182	2.8

Jonasson (2009)	Reg.	Jämföra incidensen för olika typer av cancer hos de som använder insulin glargin med de som använder andra typer av insulin (läkemedel).	Kvinnor som under 2006-2007 använde insulin glargin hade ökad incidens av bröstcancer jämfört med kvinnor som använde andra typer av insulin. Resultatet kan bero på slumpen.	114 841	32.3
-----------------	------	--	---	---------	------

\* två olika typer av magsäcksoperationer; \*\* kranskärlsoperation

individer i åldern 20-50 med ett BMI mellan 50-60, medan ett flertal olika sjukdomsgrupper ledde till exklusion. Zethelius, Gudbjörnsdottir, Eliasson, Eeg-Olofsson & Cederholms (2014) registerstudie om fysisk aktivitet är ett exempel på ytterligare en mer exkluderande studie. Studien vände sig till individer med typ 2 diabetes i ålder 30-72 år, med ett BMI > 18 och kreatinin < 150 (njurfunktionsprov). Individer med tidigare hjärtkärlsjukdom exkluderades.

Blodprovsvärden kunde användas av såväl RCTer som av registerstudier för att begränsa studiepopulationen. I registerstudierna använde man sig av vanligt förekommande blodprov såsom kreatinin (avspeglar njurfunktion). RCTerna använde sig även av prover som inte används lika frekvent i kliniken (även om det förekommer) såsom C-peptid (markör för insulinproduktion) eller antikroppar.

Inklusions- och exklusionskriterierna kan ses som ett sätt för forskare att styra hur homogen studiepopulationen i såväl RCTer som registerstudier blir. I registerstudierna finns även exempel på att kriterierna kan användas för att öka sannolikheten att de som är i interventionsgruppen också är exponerade för åtgärden som studeras. Eliasson et al (2011) är ett exempel, där de som under ett år hämtat ut minst tre recept på lipidsänkande läkemedel inkluderades i studien.

Sammanfattningsvis framkommer att de granskade RCTerna ofta utgår från en mindre urvalspopulation, till skillnad från de granskade registerstudierna där flera har som ambition att vara rikstäckande. Urvalspopulationen, speciellt för RCTerna, är inte så detaljerat beskriven. I registerstudier baseras medelvärdespatienten på en stor studiepopulation, medan medelvärdespatienten i RCTer baseras på färre individer. Det framkommer samtidigt en variation inom såväl registerstudier som RCTer gällande hur liberal man är med att använda olika exklusionskriterier. En aspekt som skiljer dem åt är att registerstudierna om de använder blodprov för exklusion, tar hjälp av vanligt förekommande blodprov. RCTerna har istället möjlighet att använda mer ovanliga blodprover då forskarna själva kan bestämma vilka prover som ska tas. För forskarna innebär RCTens studiedesign alltså en större möjlighet till kontroll.

Hur studiepopulationen blir påverkas inte bara av urvalspopulation och olika kriterierna, utan även av rekryteringsstrategier. Strategier som användes för att rekrytera deltagare var beskrivet i begränsad omfattning. För registerstudierna ser jag i första hand faktumet att 10-30% av individer med diabetes inte är med i NDR som en betydelsefull faktor. (10-30% var de siffror som de flesta registerstudierna rapporterade.) Det går inte att per automatik anta att de som är med i registret är lika de som inte är med. I RCTerna användes olika rekryteringsstrategier. Guldbrand et al (2012, s. 2126) skriver hur patienter kontaktades av sjuksköterskor för inklusion i studien. Sjuksköterskorna hade redan innan studien hand om patienterna, vilket kan ha påverkat vilka som hade möjlighet att inkluderas i studien. Detta skriver författarna kan vara en faktor som kan ha bidragit till en hög deltagarfrekvens, vilket i sin tur kan leda till att den allmänna tillämpbarheten av studieresultaten kan vara begränsat. Fritz et al (2011) använde sig av tidningsannonser samt vände sig till deltagare av en tidigare studie. De skriver att det teoretiskt kan leda till bias då en del av deltagarna varit med i en tidigare studie, men de ser det som osannolikt.

Det är dock inte alltid som de olika rekryteringsstrategiernas innebörd för den slutliga studiepopulationen kommenteras. Det kan tänkas att det påverkar studiepopulationen så att den inte speglar urvalspopulationen. Vad detta i så fall innebär när man vill överföra resultaten till andra sammanhang är svårt att säga, speciellt för de studier där även urvalspopulationen var begränsat

beskriven. I de granskade RCTerna diskuteras inte den externa validiteten i någon större utsträckning, även om undantag finns. Risstad et al (2015, s. 360) som studerat magsäcksoperationer skriver att:

The results may not generalize to patients with BMI outside the range of 50 to 60.

Ett tecken på att studiedeltagare kan skilja sig från den kliniska verkligheten framkommer i Fritz et al (2011) RCT. De jämförde studiedeltagarnas baslinjevärden för livskvalitet med motsvarande data för en ålders- och köns matchad svensk kontroll. Vid jämförelsen med den svenska kontrollen såg man t.ex. att studiedeltagarna med normal blodsockertolerans fick signifikant högre poäng på 8 av 13 områden. Det går dock inte att säga om detta beror på att studiepopulationen skiljer sig från urvalspopulationen eller om det är så att urvalspopulationen skiljer sig från den svenska kontrollgruppen eller om det beror på någon annan förklaring. I RCTerna kan olika strategierna såsom att patienterna själva aktivt har sökt studien eller att patienterna har valts ut av vårdpersonal med kännedom om dem, bidra till att de inkluderade patienterna är mer motiverad eller har högre följsamhet än de man möter vardagligen. Dessa strategier kan också verka för att patienterna är friskare i RCTer än i den kliniska verkligheten, även om det inte måste vara så.

Flera av registerstudierna lyfter å andra sidan fram extern validitet som en av deras styrkor. Holzmann et al (2015, s. 1650) skriver:

...considering the nationwide design of our study and the recent studyperiod, reflecting current standard of care, we believe that our results have a high external validity and are applicable to other patients with DM undergoing CABG in countries with a similar level of healthcare.

Den externa validiteten kopplas alltså samman med att studien är rikstäckande och speglar de förhållanden som råder i den kliniska verksamheten.

## Hur skildras patienter?

De granskade registerstudierna använder flera olika register (allt ifrån tre register upptill sju register). Register som används är dödsorsaksregistret och olika hälsodataregister, vilka Socialstyrelsen ansvarar för (Socialstyrelsen u.å.). Holzmann et al (2015) registerstudier har även valt att samla in data om socioekonomiska faktorer. För att göra det har de använt sig av statistiska centralbyråns register. Det är dock inte alla registerstudier som presenterar socioekonomiska faktorer. Gemensamt för alla de inkluderade registerstudierna är att de använt sig av NDR (Nationella Diabetesregistret). NDR är ett kvalitetsregister som används för både forskning och för att jämföra olika enheter (t.ex. vårdcentraler eller sjukhus) både med varandra och med sig själva. Flera av registerstudierna beskriver att NDR har en god täckningsgrad då 70-90% av individer med diabetes är med (Ekström et al 2012; Holzmann et al 2015; Eliasson et al 2011; Zethelius et al 2014). Jonasson et al (2009) som studerar associationen mellan en typ av insulin och cancer skriver dock att NDR bara täckte omkring hälften av de studerade individerna. RCTerna insamlar istället själva sin egen data och är därför inte på samma sätt beroende av vad andra ser som relevant data att samla in.

I bägge studietyperna återfinns långa listor på medelvärden av olika baslinje-parametrar. Medelvärden återfanns i de flesta studierna för exempelvis HbA1c (långtidsblodsocker), blodtryck och BMI. Patienter beskrivs även genom olika grupptillhörighet, exempelvis presenteras hur stor andel som röker i de flesta studierna. I registerstudierna finns icke-medicinska faktorer som inte återfinns hos RCTerna såsom andel inom ett visst civilstånd eller med en viss utbildningsnivå. RCTer har å andra sidan faktorer som kan vara svårt att få fram om man enbart använder sig av register. I RCTerna återfinns andel med sömnapné presenterat liksom andel med nedsatt glukostolerans respektive metabolt syndrom. Vid jämförelser av enskilda parametrar eller grupptillhörigheter framkommer en variation inom såväl registerstudier som RCTer. Lägst medelvärde för ålder finns i RCTen av Risstad et al (2015), där är genomsnittsåldern 35-36 år. Det högsta medelvärdet presenteras i registerstudien av Holzmann (2015), där medelåldern är 67 år. Det högsta medelvärdet för BMI återfinns i Risstads et al RCT, vilket ligger på cirka 55. Holzmann et al (2015) registerstudie och Alvarssons et al RCT

(2003) har bland de lägsta BMI-medelvärdena, vilka ligger runt 27-28. För presentation av ålder respektive BMI för de inkluderade patienterna se tabell 2 och tabell 3. Att jämföra den totala sjukdomsbördan i de olika studierna är utanför denna uppsats omfång, däremot utifrån att åldern generellt sett är något högre i de inkluderade registerstudierna tyder det på att dessa patienter kan vara sjukare än patienterna i de inkluderade RCTerna.

**Tabell 2** I tabellen presenteras registerstudiernas inklusionskriterier gällande ålder respektive BMI samt studiedeltagarnas genomsnittsålder och genomsnitts-BMI

Förste författare	Ålders-kriterier (år)	Genomsnitts-ålder (år)	BMI inklusions-kriterier (kg/m <sup>2</sup> )	Genomsnitts-BMI
Holzmann	-	67	-	27
Ekström	40 < 85	65	-	30
Eliasson	18-75	Cirka 62	-	nästan 30
Zethelius	30-72	60 respektive 59	>18	29 respektive 31
Jonasson	-	-	-	-

**Tabell 3** I tabellen presenteras RCTernas inklusionskriterier gällande ålder respektive BMI samt studiedeltagarnas genomsnittsålder och genomsnitts-BMI

Förste författare	Ålders-kriterier (år)	Genomsnitts-ålder (år)	BMI inklusions-kriterier (kg/m <sup>2</sup> )	Genomsnitts-BMI (kg/m <sup>2</sup> )
Guldbrand	-	61 respektive 62	-	33
Fritz	45-69	61	> 25	29
Alvarsson	35-70	51 respektive 56	<35	27 respektive 28
Chiasson	40-70	55	25-40	31
Risstad	20-50	35 respektive 36	50-60	55

I både RCTer och registerstudier är det alltså vanligt att patienter beskrivs i form av medelvärden och andelar. Det är däremot sällan som enskilda individer beskrivs, och när det förekommer är det mycket kortfattat och individens kontext är svår att skönja. I Risstad et al (2015, s. 356) artikel om magsäcksoperation återfinns en av de längre beskrivningarna av en enskild individ:

Four patients with duodenal switch (13.8%) required hospital admissions owing to protein-calorie malnutrition. One of these patients also had severe diarrhea and was subsequently operated on with elongation of the common channel 3 years after the initial procedure.

Några av studierna presenterade däremot resultat för subgrupper t.ex. för olika grader av njurfunktion (Ekström et al, 2012). Subgruppsanalys kan ses som ett försök att ta medelvärdespatienten ett steg närmre den enskilda individen i en klinisk praktik.

## Design av resultat

Vid granskning av de inkluderade RCTerna framkom en stor variation sett till vad man vill undersöka. Fokus spannar mellan betacellsfunktion till hälsorelaterad livskvalitet. Metabol kontroll och hjärtkärlsjukdom (eller risk för sådan) är ett återkommande tema i flera av studierna. I RCTerna presenteras ett flertal utfallsmått av olika karaktär. I RCTerna återfinns blodprover som rutinmässigt kontrolleras i klinisk praxis såsom HbA1c (långtidsblodsocker) och kolesterol, men det finns även blodprov som är mer ovanliga såsom glukagonstimulerad c-peptid. Förutom olika mätvärden finns

även händelser av hjärtkärlsjukdom som utfall (Guldbrand et al 2012; Alvarsson et al 2003; Chiasson et al 2003; Risstad et al 2015; Fritz et al 2011).

I de granskade registerstudierna återfinns också en stor variation sett till vad man ämnar belysa. Eliasson et al (2012) fokuserar på kolesterolnivåer (LDL-C) medan Holzmann et al (2015) är intresserad av långtidsöverlevnad efter CABG-operationer. Registerstudierna har ett färre antal utfallsmått jämfört med RCTerna. Fokus är i första hand på total dödlighet, samt dödlighet och sjuklighet kopplat till hjärtkärlsjukdom (Ekström et al 2012; Jonasson et al 2009; Holzmann et al 2015; Zethelius et al 2014). Eliasson et al (2011) artikel sticker däremot ut från gruppen då de har olika typer av kolesterolvärden som utfallsmått.

I de båda studieformerna undersöks såväl positiva som negativa effekter. Några av registerstudierna lyfter upp att det just är effekter i den kliniska praxisen man syftar till att undersöka. Begreppet effectiveness återkommer hos såväl Ekström et al (2012) som Eliasson et al (2011). Intressant nog studeras livskvalitet som utfallsmått i tre RCTer (Risstad et al 2015; Alvarsson et al 2003; Fritz et al 2011) medan det inte förekommer i någon av registerstudierna.

Även om registerstudierna ofta har färre antal utfallsmått, kunde det ändå i slutändan bli många mätvärden som presenteras. Registerstudier justerar resultaten efter olika variabler. Flera av studierna använder flera olika justeringsmodeller, men det behöver inte leda till att resultatets osäkerhet tas bort. Zethelius et al (2014) lyfter fram i deras registerstudie att det ändå finns en osäkerhet kring resultatet, då det inte går att justera för okända faktorer. Eliasson et al (2011) är ett annat exempel som synliggör en osäkerhet. De skriver att resultatet ska tolkas med viss försiktighet. Detta grundar de på att det finns skillnader mellan interventionsgrupp och kontrollgrupp, vilket pekar på en möjlig selektionsbias. De skriver att helst ska studiens resultat konfirmeras av andra prospektiva studier. En osäkerhet kopplat till resultatets innebörd återfinns även i RCTen av Chiasson et al (2003). De skriver att deras studie är hypotesgenererande, eftersom de undersöker en fråga som studien inte var designad att studera från början. Chiasson et al skriver att de själva tror att deras observationer är signifikanta statistiskt såväl som kliniskt, men att deras resultat behöver konfirmeras av andra prospektiva studier.

Hur de olika utfallsmåtten sedan används i studiernas konklusioner uppvisar som tidigare nämnts en stor spännvidd. Ekström et al (2012) kommer i sin registerstudie fram till att i klinisk praxis överväger fördelarna med metformin dess risker, detta då deras studie synliggjorde att det inte var högre risker förknippade med metformin jämfört med andra blodsockersänkande läkemedel. Holzmann et al (2015) kommer i sin registerstudie fram till att CABG-operation är förknippade med ökade risker för individer med typ 1 diabetes. Deras rekommendation blir dock inte att vara mer restriktiv med operationen, utan förordar istället en noggrannare uppföljning för dessa individer efter operationen. Holzmann et al (2015, s 1651) skriver:

...and that all possible measures to mitigate their risk of death or recurrent cardiovascular events should be instituted.

Jonasson et al (2009) registerstudie presenteras en statistiskt signifikant relativ risk (RR=1.99, 95% KI 1.31–3.03) för bröstcancer hos de som använt insulin glargin. De lyfter istället fram slumpmässiga fluktuationer som en trolig förklaring till resultatet. Detta efter sammanvägning med resultaten för de andra typerna av cancer som studerats, kombinerat med kunskap från andra källor. Det är alltså inte bara värdena på utfallsmåtten som är av betydelse för vilka slutsatser som dras i de olika registerstudierna. I Alvarsson et al (2003) RCT är tidig insulinbehandling förknippad med bättre värden för blodsocker och c-peptid, jämfört med de som fått tablettbehandling. De skriver att deras RCT ger argument för att börja med insulinbehandling i ett tidigt skede av diabetes-sjukdomen. Risstad et al (2015) jämför två olika operationsmetoder, duodenal switch respektive gastric bypass. Den först nämnda operationsmetoden medförde större viktminskning samtidigt som metoden även innebar större risk för komplikationer. Risstad et al skriver att man därför utifrån resultatet från deras RCT ska vara försiktig och tänka igenom vilka som ska opereras med duodenal switch.

Sammanfattningsvis förekommer olika typer av utfallsmått (från olika blodprover till sjukdom och död) i så väl registerstudier som i RCTer, däremot återfanns livskvalitet som utfallsmått enbart i några av RCTerna. Det kan finnas en osäkerhet sett till vad resultatet står för både i registerstudier och i RCTer. Slutsatserna i RCTerna ses oftast vara samstämmiga med utfallsmåtten. I registerstudierna finns det istället en större variation sett till hur slutsatsen relateras till studiens utfallsmått.

## 6 Diskussion

I detta avsnitt börjar jag med en genomgång av olika sätt att se på begreppsparet stringens-relevans som framkommit i uppsatsen (6.1). Därefter kommer design av patient och resultat i RCTer respektive registerstudiernas att vara i fokus (6.2), för att sedan återkomma till begreppen stringens och relevans (6.3 och 6.4), men i blickpunkt nu är olika perspektiv som de enskilda begreppen kan förstås utifrån. Avslutningsvis återkommer jag till induktionsproblemet som presenterades i början av uppsatsen (6.5). Hur hanterar RCTer respektive registerstudier induktionsproblemet?

### 6. 1 Begreppsparet stringens-relevans

Vid genomgång av studieformernas historiska ursprung och de samtida beskrivningarna av vilken typ av kunskap som RCTer respektive registerstudier kan bidra med, framträder spänningen mellan relevans och stringens. RCTen lyfts fram av flera författare som en studiedesign som ökar resultatets säkerhet. RCTen framhåller man, har potential att minska inflytandet av confounders och på så sätt leda oss rätt i valet av behandling. Fries (1976) ger samtidigt uttryck för att den kliniska relevansen av resultaten kan ifrågasättas. Han skriver att de frågor man har för stunden sällan har studerats, där prospektiva studier ofta har brister såsom att de är för tidsbegränsade eller för exkluderande. Det kan ses som en kontrast till registerstudier, vilka kopplas samman med bredare fokus på olika utfall. Man menar att registerstudien öppnar upp en möjlighet till att få större klarhet i beslut för individer som är underrepresenterade i RCTer. Registerstudierna ses kunna svara på relevanta frågor vilket Wennbergs utlåtande kan vara ett exempel på. Han lyfter upp att den kunskap som är mest relevant för läkarna i Maine finns i register över utfallen i Maine, men även om registerstudier ses kunna svara på relevanta frågor så framkommer en osäkerhet sett till vilken trovärdighet dessa svar har. Både Byar (1980) och Taube (2008) ger exempel på hur registerstudien inte ger någon vägledning om vad dess resultat står för. Det finns istället flera olika sätt att tolka resultatet på, vilket också är tydligt i den granskade studien av Jonasson et al (2009). De studerar associationen mellan insulin glargin och bröstcancer. Trots en stor relativ risk (RR) för bröstcancer hos de som fått förskrivit insulin glargin, skriver de att det inte går att säga vad detta resultat står för – slumpen eller att insulin glargin kan leda till bröstcancer. En spänning mellan metodens stringens och resultatets relevans målas upp.

Ett sätt att tolka valet mellan stringens och relevans är att det handlar om ett val mellan behovet av kontroll och behovet av att avspegla den kliniska vardagliga praxisen. Grossman och Mackenzie (2005) som i sin artikel ifrågasätter värdet av randomisering synliggör samtidigt att RCTer kan förknippas med behovet av kontroll. De pekar ut de kliniska fas I-III undersökningar som exempel där man är i behov av metodens möjlighet till kontroll, eftersom man vid dessa undersökningar studerar nya läkemedel. Valet mellan behovet av kontroll och behovet av att avspegla den kliniska praxisen återfinns i den RCT som jämför två olika dieter. Artikelns författare (Guldbrand et al, 2012) skriver att i slutet av deras studieperiod hade flera deltagare börjat med kolesterolsänkande läkemedel, vilket begränsar studiens interna validitet. Valet att inte styra vad som hände deltagarna i övrigt var en konsekvens av deras försök att få studien så klinisknära som möjligt. Den kliniska ambitionen kan kontrasteras mot ett behov av kontroll. Ju striktare kontroll över vilka individer som är med i studien, hur åtgärderna ser ut samt vilka andra åtgärder som är tillgängliga, desto lättare ter det sig att renodla resultatet. Den enda skillnaden som finns mellan de två grupperna är den intervention man vill undersöka, vilket Taube (2008) lyfter upp i sin artikel. I ett utslag av stort kontrollbehov så lyfter Jernberg och Janzon (2015) samt Norlin et al (2014) fram risken att man inte befinner sig nära den kliniska verkligheten. Patienterna som inkluderats i studien kan vara friskare eller ha en orealistisk följsamhet jämfört med de patienter som man möter i den kliniska praktiken.

Frågan är ifall stringens och relevans behöver ses som ett motsatspar? RRCT kan vid en första anblick ses som ett uttryck för att vilja kombinera stringens och relevans genom att få tillgång till både registerstudiens styrka i form av närhet till den kliniska praktiken samt RCTens säkerhet genom att randomisera deltagare. När RRCTen granskas närmare framkommer en bild av att det snarare handlar

om en fortsatt balansgång mellan stringens och relevans, men att balanseringen av relevans och stringens är annorlunda jämfört med hur det sker i registerstudier respektive RCTer. Metoden medför inte tillräcklig kontroll för att kunna användas i kliniska fas I-III studier, samtidigt som de som var mest sjuka (och kanske de som var i mest behov av ett tydligt svar?) inte fick delta i själva studien. Walach et al (2006) lyfter i sin artikel upp dilemmat med att det inte med studiedesign går att lösa metodens tillkortakommande. Istället skriver författarna att det får göras efter att studierna är gjorda, där cirkelmodellen är ett tillvägagångssätt. Modellen innebär att olika studiers styrkor och svagheter kombineras, vilket kan ses som ett uttryck för att vilja kombinera både relevans och stringens. Cirkelmodellen lyfter fram att olika frågor bäst besvaras genom olika typer av studiedesign, men även detta kan uppfattas ha ett pris, nämligen en avsaknad av en tydlig vägledning. Ifall resultaten pekar åt olika håll, finns ingen hierarki som lyfter fram vilka resultat som ska prioriteras. Istället blir det troligen en bedömning om studiernas kvalitet och om metoderna varit lämpade att besvara de frågorna som ställts, som får avgöra vilka studier som ska premieras på bekostnad av andra. Dilemmat med att både inkludera olika studieformat och samtidigt ha en tydlig vägledning kan ses hanteras på ett lite annorlunda sätt i GRADE. I GRADE rangordnas RCTer högre än observationsstudier vid den preliminära bedömningen av evidensstyrkan. Olika faktorer såsom överförbarhet, effektstorlek, precision kan sedan både sänka och höja evidensstyrkan. I GRADE finns alltså hierarkin kvar vid evidensgradering, men den slutliga evidensstyrkan från observationsstudier behöver alltså inte bedömas som lägre än evidensstyrkan från RCTer. Frågan är om detta är tillräckligt för att eliminera Howicks (2011) farhåga? Howick lyfter ju som tidigare beskrivet upp risken med att ett hierarkiskt system kan leda till att observationsstudier ignoreras när det finns RCTer.

## 6. 2 Design av patient och resultat

Både i registerstudier och i RCTer beskrivs de inkluderade patienterna i form av en aggregerad medelvärdes patient som är kontextlös. Bilden som framkommer efter en närmare granskning av litteraturen är dock att patienterna i RCTer ofta är yngre, har högre följsamhet och är friskare jämfört med de patienter som man möter i den kliniska verkligheten. RCTer kan dessutom bestå av en lägre andel kvinnor. Det kan förklaras genom snäva inklusionskriterier, även om det finns röster, däribland Howick (2011), som menar att RCTernas inklusionskriterier inte behöver vara snäva. Att RCTer inte behöver ha snäva inklusionskriterier syns vid granskningen av exempelvis Guldbrands et als (2012) RCT. Patienter i registerstudier ses istället som en spegling av den kliniska verkligheten. Ibland beskrivs registerstudier även i termer av att studierna *är* den kliniska verkligheten, vilket syns i Jernberg och Janzon (2015) artikel. De skriver att i kliniska prövningar var andelen med kardiovaskulär död ungefär hälften av det vi ser i den kliniska verkligheten, där den kliniska verkligheten bestod av data från en registerstudie. En förklaring till att registerstudier ses om en spegling av den kliniska praxisen beskriver Jernberg och Janzon (2015) är att de inte är lika exkluderande som RCTer. Författare såsom Byar (1980) och Norlin et al (2014) synliggör att det i registerstudier finns en osäkerhet sett till vem patienten egentligen är, om den tar det ordinerade läkemedlet och om ordination är på rätt grunder? Att patienter inte selekteras bort och att det i registerstudier precis som i den kliniska verkligheten föreligger en osäkerhet, kan ses som att studien är närmare den kliniska praktiken. Detta synliggör ett sätt att se på begreppet relevans, nämligen egenskapen att vara nära den kliniska praktiken.

Även när resultaten från registerstudier och RCTer granskas finns en gemensam nämnare, nämligen att det är svårt att direkt applicera resultatet på den enskilda individen. Resultatet i registerstudier präglas av en osäkerhet sett till vad det egentligen står för, vilket grundas på flera olika faktorer. Förutom att det råder en osäkerhet till vem det är som studerats medför frånvaron av randomisering att någon (antingen patient eller vårdpersonal) valt vilka som ska få en viss intervention. Det möjliggör att det kan finnas andra gemensamma faktorer som karakteriserar de patienter som har fått interventionen. Det kan alltså förutom interventionen finnas andra viktiga skillnader mellan interventionsgrupp och kontrollgrupp som förklarar resultatet. Det är särskilt viktigt att ha i beaktande vid studier av icke-akuta tillstånd. Det lyfts därför fram som extra viktigt för resultatet trovärdighet att i registerstudier



justera för socioekonomiska faktorer. Intressant nog så var det inte alla granskade registerstudier som reflekterade över just detta. Osäkerheten som återfinns i registerstudier kan ses som att det gör att studien är närmare den kliniska praxisen än en hårt styrd RCT, och därmed är mer relevant. Frågan man kan ställa sig är om den osäkerhet som finns i registerstudierna bidrar till att minska osäkerheten när ett kliniskt beslut ska tas?

Det finns samtidigt olika faktorer som beskrivs minska registerstudiens osäkerhet, även om det råder delade meningar om det leder till ett trovärdigt resultat. Faktorer som lyfts upp är att studierna använder sig av trovärdig data, att det finns god förståelse för prognostiska faktorer samt användandet av statistisk teknik. Andra aspekter att ha i beaktande är effektstorleken i förhållande till eventuella confounders och studiens kvalitet.

Resultaten från RCTer beskrivs av författare såsom Taube (2008) och Howick (2011) genom studiedesignen som trovärdiga. RCTen möjliggör att medicinskt viktiga men små skillnader framträder. Det beror på att studiepopulationen är homogen både som helhet men också när interventions- och kontrollgrupp jämförs. Genom att RCTen betraktas som en stringent metod, kan det finnas en risk att man inte ifrågasätter hur stringent RCTen egentligen är. Har randomiseringen skapat de homogena grupper man tänk? Är rätt enhet randomiserad? Det framkommer dessutom vid litteraturgenomgången att det inte alltid är relevant att randomisera, exempelvis kan möjligheten att randomisera deltagare begränsas av etiska aspekter.

### 6.3 Olika aspekter av begreppet relevans

Vid granskning av design av patient och resultat i RCTer respektive registerstudier så framkommer att begreppet relevans kan ses utifrån olika perspektiv. Precis som i Knaapens (2013) artikel kan relevans kopplas samman med om studiens resultat är överförbart till en annan kontext respektive om studien kan bidra med svar till frågan. Det framkommer även andra aspekter av begreppet relevans. Relevans kan förknippas med att studien är nära den kliniska praxisen. Denna aspekt ses ofta som synonym med hög överförbarhet, vilket exemplifieras av att registerstudier lyfts fram av olika författare som spegling av den kliniska praxisen och att resultatet därför ses vara överförbart till den kliniska verkligheten. Frågan är dock om spegling av den kliniska praxisen alltid ska ses som synonymt med att resultatet är överförbart? De granskade registerstudierna som var baserade på NDR, beskriver att 10-30% av individer med diabetes inte är med i registret. Vad innebär denna aspekt när resultat från registerstudier ska överföras till den kliniska praktiken? Kan man förvänta sig att de resultat som finns i en registerstudie, förutsatt att de har hög trovärdighet, även går att uppnå i den kliniska praktiken? Det finns även en obesvarad fråga sett till vad det innebär att basera forskning på registerdata som även används för att jämföra olika enheter med varandra. Thörn och Järhult (2015) lyfter upp farhågan om att data snedvrids eller t.o.m. manipuleras.

Förutom de ovannämnda perspektiven kan relevans även ses ur ett patientnära perspektiv, där relevans kopplas samman med kunskap som är av direkt värde för den enskilda patienten. Utfall som fokuserar på funktionsnivå och patientnöjdhet kan ses vara av stort värde för den enskilda patienten. Det är inte bara av intresse att veta om man kommer att leva eller dö, utan även hur man kommer må och vad man kommer kunna göra, som är av värde. Intressant nog var dessa utfall något som under utvecklingen av outcomes research just kopplades till registerstudier. I de granskade studierna är det flera RCTer som studerar livskvalitet, samtidigt som ingen av registerstudierna har detta fokus. Hur nära studieresultatet uppfattas vara en enskild individ kan förstås på olika sätt. Registerstudier kan bestå av mycket stora populationer (över 100 000), där studien är av inkluderande karaktär. Detta kan ses leda till att det är mer troligt för den enskilda patienten att den har ingått i studien/skulle kunna ha ingått i studien, samtidigt som det fortfarande är oklart för den enskilda patienten om just hen hade haft vinning av interventionen. Ifall interventionen är till gagn för den enskilda patienten kan göras mer eller mindre troligt genom att dela in studiepopulationen i olika subgrupper. Subgruppsanalyser kan vara ett uttryck för att ta studiens medelvärdespatient och dess resultat ett steg närmre patienten i den kliniska verkligheten. Genom snäva inklusionskriterier kan RCTens medelvärdespatient ses vara

längre ifrån den kliniska verklighetens patienter. En RCT med snävare inklusionskriterier som består av ett hundratal individer kan samtidigt ses vara närmare den enskilda individen, förutsatt att individen matchar inklusions- och exklusionskriterierna.

Det framkommer alltså flera olika sätten att se på relevans. Att ge ett trovärdigt svar eller att resultatet är överförbart kan ses som relevant. Relevans kan även förstås utifrån att avspegla den kliniska praxisen eller vara patientnära. Även om avspeglingen av den kliniska praxisen ofta ses som synonymt med hög överförbarhet, behöver det inte vara så för varje enskild studie. Ett sätt att hantera de olika sätten att förstå relevans är att tydliggöra vilka typer av relevans som man ser att en studie bidrar med i förhållande till vad man i det kliniska beslutet är i behov av.

## 6.4 Olika sätt att se på stringens

Vid granskningen av stringens framkommer inte lika stor variation som för begreppet relevans, men två sätt att se på stringens framträder. Tydligast är hur olika typer av studiedesign rangordnas efter metodens stringens i en hierarkisk modell. I evidenspyramiden placeras RCTer över registerstudier. Stringensbedömningen sker i detta fall mellan olika metoder. Det framkommer även tankegångar om att det viktigaste inte är att rangordna studier efter den använda metodens potentiella stringens. Det viktiga är istället att se ifall forskarna använder metoden på ett stringent sätt, vilket leder till en betoning av vikten av kvalitetsgranskning och att lämplig metod använts för att besvara frågan. Kvalitetsgranskning beskrivs av författare såsom Rosén et al (2008) som viktigt för såväl registerstudier som för RCTer. Stringensbedömningen sker i detta fall snarare inom själva metoden än mellan olika metoder. Det kan få till följd att det inte alltid är RCTen som ses som det mest stringenta alternativet.

## 6.5 Induktionsproblemet

I både RCTer och registerstudier presenteras en patient som inte finns - en generell, universell patient. Ett sätt att se på denna process är att man utifrån få observationer gör en generell utsaga. Det leder oss fram till induktionsproblemet, nämligen att ett begränsat antal observationer aldrig kan producera en universell utsaga. Det går inte med säkerhet att veta att det mönster som skapats utifrån  $n$  antal observationer även gäller observation  $n+1$ . Resultatet som fås från en RCT eller registerstudie kan alltså inte med visshet gälla den patienten som man möter i den kliniska praktiken. Det kan samtidigt vara problematiskt att kritisera RCTer utifrån induktionsproblemet. RCTen (åtminstone i sitt idealfall) ses som en deduktiv metod. Ser man på RCTen som en metod som har likheter med ett falsifieringsperspektiv, gör RCTen inte anspråk på att komma fram till en universell sanning. Studien syftar inte till att producera en universell utsaga utan kan istället ses som att stegvis förbättra teorin. Detta sker genom att testa hypoteser och förkasta de teorier som falsifierats, samtidigt som falsifiering inte är problemfritt. Duhem-Quine-tesen lyfter upp att en teori aldrig kan testas isolerat, utan den testas tillsammans med olika antaganden och förutsättningar. Så ifall hypotesen som testas i en RCT falsifieras, behöver det inte vara teorin som är fel. Det kan även vara såsom Sismondo (2010) skriver att det är mätapparaturen som används som inte är tillförlitlig, eller att okända objekt och processer påverkar observationerna. RCTens strikta kontroll och randomisering kan ses som ett sätt att minska risken för att det är andra aspekter än själva teorin som testas, även om det inte går att undkomma. Duhem-Quine-tesen kan ses som en ännu större utmaning för registerstudier, då de har mindre kontroll över vad som egentligen testas. Detta kan ses leda till att det är svårare att tolka resultatet, då det kan råda större oklarhet vad ramverket som testas består av. Utifrån detta perspektiv kan det därför ses som att RCTer och registerstudier hanterar induktionsproblemet på liknande sätt – nämligen genom att inte utgå från en induktiv metod.

Det går också att se det som att RCTer och registerstudier hanterar induktionsproblemet på olika sätt. Taube (2008) skriver att registerstudier inte är hypotesprövande utan snarare är hypotesgenererande, vilket kan ses som att registerstudier har likheter med en induktiv metod. Sätt att hantera induktionsproblemet i registerstudier kan då ses vara att bygga sina studier på ett stort antal

observationer (i vissa fall över 100 000 deltagare). Chalmers (2013) skriver belyser dock att det är oklart hur många observationer som behövs för att de ska ses som tillräckligt många. Förutom ett stort antal observationer kan registerstudier ses hantera induktionsproblemet genom att de i studien ( $n$  antal) beskrivs vara lika individen utanför studien ( $n+1$ ). Sismondo belyser samtidigt att bedömning av *sameness*, inte är lätt. Ofta finns inte ett väldefinierat koncept för vad *sameness* är, vilket innebär att det finns flera olika tankegångar om vad det innebär att fallet  $n+1$  är samma som det första  $n$ . Oavsett om studiers metod har likheter med en induktiv metod eller inte, är det ett intressant dilemma som Sismondo belyser, när man funderar på om ett resultat från en studie går att överföra till en annan kontext. Hur lik är studien den kontext som jag vill överföra resultatet till? Hur lik måste den vara för att jag ska kunna överföra resultatet? Skiljer sig sättet att se på behovet av *sameness* när resultat från registerstudier respektive RCTer ska överföras till den kliniska praktiken? Det kompliceras ytterligare när Cartwright och Hardie (2012) tas i beaktan. För att bedöma om resultatet från en studie går att överföras till den kliniska praxisen räcker det inte med att bedöma hur lik studien är den kliniska praxisen. Det behövs specifikt fokus på om nödvändiga stödfaktorer finns närvarande i den kliniska praktiken och om åtgärden har potential att utgöra samma orsaksroll i den kliniska praxisen som den gjorde i studien.

## 7 Epilog

Denna uppsats började med en beskrivning av de två fiktiva patienterna Pernilla och Tibor, vilka vi nu återvänder till. Det är dags att hantera det dilemma som induktionsproblemet synliggjort - nämligen överföring av kunskap från olika studiers generella medelvärdes patienter till två enskilda patienter i den kliniska vardagen. Så vad innebär denna uppsats resultat och diskussion sett till vilken hjälp jag kan tänkas få från RCTer respektive registerstudier, när beslutet ska tas om Pernilla och Tibor ska ordinerats någon av de nya blodsockersänkande medicinerna? Vilken typ av kunskap om ett läkemedels effekt kan RCTer respektive registerstudier bidra med till det kliniska beslutet?

I litteraturen framkommer tre förhållningssätt för hur man ser på vilken kunskap om ett läkemedels effekt som kan fås från registerstudier respektive RCTer. Det första är att RCTen är den studieform som ger säkrast kunskap sett till behandlingseffekter och därför ska användas, samtidigt som registerstudier ska undvikas. Det andra förhållningssättet är att RCTer bidrar med säkrare kunskap men att registerstudier kan vara till hjälp i de fall det saknas RCTer. Det tredje alternativet är att RCTer och registerstudier ses bidra med olika typer av kunskap, där RCTen inte alltid är att föredra. Det första förhållningssättet kan ses som ett uttryck för betoning av stringens, medan det tredje alternativet kan ses som en betoning av relevans. Skillnaden mellan alternativ ett och alternativ tre går också att förstå utifrån olika tolkningar av stringens. Det första alternativet premierar stringensbedömning som görs mellan metoder, medan det tredje alternativet premierar stringensbedömning inom metoden.

Då det handlar om nya läkemedel så har läkemedlen studerats genom olika RCTer för att kunna bli godkända. Utifrån att Pernilla troligen passar inklusions- och exklusionskriterierna sett till både ålder, sjukdomsburda och BMI, kan RCTer ses bidra med viktigt kunskap om förväntade utfall och vanliga risker och biverkningar. För Pernilla kan RCTer ses bidra med ett trovärdigt svar. Även om RCTerna inte ses spegla den kliniska praxisen, kan de ändå i Pernillas fall ses vara patientnära sett till att hon är lik de som inkluderats i studierna. Frågan är dock om hon är tillräckligt lik? Det kan finnas frågetecken om hennes följsamhet motsvarar den som återfinns i RCTerna, detta då det i den kliniska verkligheten är svårt att uppnå samma kontroll som kan finnas i RCTer.

För Pernilla kan registerstudier tänkas bidra med kunskap om långtidsrisker eller ovanliga risker. Dock som det syns både i litteraturgenomgången och Jonassons registerstudie (2009), kan det vara svårt att bedöma betydelsen av risker som framkommer i registerstudier men inte syns i RCTer. Å ena sidan kan registerstudier ha större möjlighet att synliggöra riskerna. Å andra sida kan det vara svårare att tolka dess trovärdighet. Vid bedömning av registerstudiens trovärdighet är en viktig faktor om studien vägt in socioekonomiska faktorer.

Kunskap om risker från registerstudier kan även bidra med relevant kunskap för Tibor. I mötet med Tibor är man kanske mer återhållsam med att ordinera nya läkemedel ifall det framkommer misstanke om allvarliga risker. Detta med tanke på att Tibor har en större sjukdomsburda i form av tidigare hjärtinfarkter samt njursvikt. Det finns samtidigt en potential att Pernilla, sett till att hon är yngre, kommer exponeras för läkemedlet under en längre tid, vilket också är bra att ta i beaktan när risker ska försöka värderas.

För Tibor är det så klart viktigt att ta hänsyn till resultaten från RCTerna. Det kan ha framkommit aspekter som både kan stärka läkemedlets position eller som kan leda till att det inte bedöms vara aktuellt. Det kan dock råda större frågetecken om resultaten i RCTerna även är giltiga för Tibor. Det kan finnas restriktioner kopplat till RCTerna, sett till att de inte har inkluderat individer som har nedsatt njurfunktion som är i nivå med Tibors. Registerstudier kan då ses bidra med klinisk erfarenhet som är större än ens egna kliniska erfarenhet. I registerstudier kan man se hur utfallen är för de som är ordinerade läkemedlen, med avseende både på sjuklighet och dödlighet, vilket kan ha gjorts i ljuset av olika nivåer av njurfunktion. Det finns samtidigt en osäkerhet sett till om resultatet i första hand

speglar vad som sker i den kliniska verkligheten eller om det också bidrar med kunskap om läkemedlets effekter. För att kunna bedöma resultatets trovärdighet finns det, förutom socioekonomiska faktorer, olika aspekter som är bra att känna till. Vilken kvalitet har registerstudien? Hur ser balansen ut mellan interventionens uppmätta effekt och confounders påverkan på resultatet?

Utifrån ovanstående resonemang bedömer jag att tredje alternativet är adekvat för både Pernilla och Tibor, då både RCTer och registerstudier kan bidra med kunskap till det kliniska beslutet. RCTer har ett gemensamt drag med registerstudier, nämligen att de inte till 100% kan garantera ett specifikt utfall för varken Tibor eller Pernilla. De olika metoderna kan istället ses bidra till den sannolikhetsbedömning som görs där fokus är på vad Tibor respektive Pernilla kommer gynnas mest av. Antingen om det är att ordineras någon av de nyare medicinerna eller om det är att vi väljer ett annat alternativ.

Efter dessa två fiktiva patientbesök väcks nya frågor. I den kliniska praktiken, i en medicinsk kontext, är det skillnad på hur rekommendationer baserade på RCTer respektive RCTer och registerstudier betraktas? Hur upplevs deras trovärdighet respektive användbarhet? Det finns även funderingar på vad den nya studiedesignen RRCT kommer innebära för framtida kliniska beslut. Kommer denna studieform premieras på bekostnad av registerstudier och/eller RCTer? Eller kommer de orosmoln som finns nu, i form av etiska problem och osäkerhet till datakvaliteten, växa sig större?

## Referenslista

- Alvarsson, M., Sundkvist, G., Lager, I., Henricsson, M., Berntorp, K., Fernqvist-Forbes, E., Steen, L., Westermarck, G., Westermarck, P., Örn, T. & Grill, V. (2003). Beneficial Effects of Insulin Versus Sulphonylurea on Insulin Secretion and Metabolic Control in Recently Diagnosed Type 2 Diabetic Patients. *Diabetes Care*, 26:8.
- Andréasson, S., Chikritzhs, T., Dangardt, F., Holder, H., Naimi, T. & Stockwell, T. (2016). Måttlig alkoholkonsumtion ger ingen positiv hälsoeffekt. En kritisk forskningsanalys. *Läkartidningen*, 113. <http://www.lakartidningen.se/Klinik-och-vetenskap/Klinisk-oversikt/2016/02/Mattlig-alkoholkonsumtion-ger-ingen-positiv-halsoeffekt/> [2016-05-12]
- Bohlin, I. (2011). Evidensbaserat beslutsfattande i ett vetenskapsbaserat samhälle. Om evidensrörelsens ursprung, utbredning och gränser. I Bohlin, I. & Sager, M. (red). *Evidensens många ansikten*. Lund: Arkiv förlag.
- Bohlin, I. & Sager, M. (2011). Inledning. Evidensbaserad praktik i praktiken. I Bohlin, I. & Sager, M. (red). *Evidensens många ansikten*. Lund: Arkiv förlag.
- Bradford-Hill, A. (1990). Suspended judgment. Memories of the British Streptomycin Trial in Tuberculosis. The First Randomized Clinical Trial. *Controlled Clinical Trials* 11..
- Byar, D.P. (1980). Why data bases should not replace randomized clinical trials. *Biometrics* 36.
- Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties*, 2:1.
- Cartwright, N. & Hardie, J. (2012). *Evidence-Based Policy A Practical Guide to Doing It Better*. Oxford: Oxford University Press.
- Chalmers, A. (2013). *What is this thing called science?* 4ed, Berkshire: Open University Press.
- Chiasson, J.-L., Josse, R.G., Gomis, R., Hanefeld, M., Karasik, A., & Laakso, M. (2003). Acarbose Treatment and the Risk of Cardiovascular Disease and Hypertension in Patients With Impaired Glucose Tolerance. The STOP-NIDDM Trial. *JAMA*, 290:4.
- Concato, J., Shah, N. & Horwitz, R.I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *The New England Journal of Medicine*, 342:25.
- Ekström, N., Schiöler, L., Svensson, A-M., Eeg-Olofsson, K., Miao Jonasson, J., Zethelius, B., Cederholm, J., Eliasson, B., Gudbjörnsdóttir, S. (2012). Effectiveness and safety of metformin in 51 675 patients with type 2 diabetes and different levels of renal function: a cohort study from the Swedish National Diabetes Register. *BMJ Open*, 2:e001076.
- Eliasson, B., Svensson, A-M., Miftaraj, M., Miao Jonasson, J., Eeg-Olofsson, K., Andersson Sundell, K. & Gudbjörnsdóttir, S. (2011). Clinical Use and Effectiveness of Lipid Lowering Therapies in Diabetes Mellitus – An observational Study from the Swedish National Diabetes Register. *PLoS ONE*, 6:4.
- Epstein, A. M. (1990). The outcomes movement - will it get us where we want to go? *The New England Journal of Medicine* 323.

Evidence-Based Medicine Working Group (1992). Evidence-Based Medicine A New Approach to Teaching the Practice of Medicine. *JAMA*, 268:17.

Fagrell, B. (2016). Åter igen en provokativ artikel om alkohol från Andréasson och medarbetare! [Läkartidningens kommentarsfält], 2 februari. <http://www.lakartidningen.se/Klinik-och-vetenskap/Klinisk-oversikt/2016/02/Mattlig-alkoholkonsumtion-ger-ingen-positiv-halsoeffekt/> [2016-05-12]

Fries, J.F. (1976). A databank for the clinician? *The New England Journal of Medicine* 29.

Fritz, T., Caidahl, K., Osler, M., Östenson, C. G., Zierath, J. R. & Wändell, P. (2011). Effects of Nordic walking on health-related quality of life in overweight individuals with Type 2 diabetes mellitus, impaired or normal glucose tolerance. *Diabetic Medicine*, 28.

Fröbert, O. & Cizinsky, S. (2014). Etiska kommittéer kan vara oetiska. Evidensbaserad medicin i fara. *Läkartidningen*, 111:47:C7FX. <http://www.lakartidningen.se/Opinion/Debatt/2014/11/Evidensbaserad-medicin-i-fara/> [2016-05-12]

Fröbert, O., Lagerqvist, B., Olivecrona, G.K., Omerovic, E., Gudnason, T., Maeng, M., Aasa, M., Angerås, O., Calais, F., Danielewicz, M., Erlinge, D., Hellsten, L., Jensen, U., Johansson, A. C., Kåregren, A., Nilsson, J., Robertsson, L., Sandhall, L., Sjögren, I., Östlund, O., Harnek, J. & James, S.K. (2013). Thrombus Aspiration during ST-Segment Elevation Myocardial Infarction. *The New England Journal of Medicine*, 369:17.

Gerszten, P. (1998). Outcomes Research: A Review. *Neurosurgery* 43(5).

Grossman, J. & Mackenzie, F. J. (2005). The Randomized Controlled Trial: gold standard or merely standard? *Perspectives in Biology and Medicine*, 48:4.

Guldbrand, H., Dizdar, B., Bunjaku, B., Lindström, T., Bachrach-Lindström, M., Fredriksson, M., Östgren, C.J. & Nystrom, F.H. (2012). In type 2 diabetes, randomisation to advice to follow a low-carbohydrate diet transiently improves glycaemic control compared with advice to follow a low-fat diet producing a similar weight loss. *Diabetologia*, 55.

Hlatky, M. (1991). Using databases to evaluate therapy. *Statistics in Medicine*, 10.

Holzmann, M.J., Rathman, B., Eliasson, B., Kuhl, J., Svensson, A.-M., Nyström, T. & Sartipy, U. (2015). Long-Term Prognosis in Patients With Type 1 and 2 Diabetes Mellitus After Coronary Artery Bypass Grafting. *Journal of the American College of Cardiology*, 65:16.

Howick, J. (2011). *The Philosophy of Evidence-Based Medicine*. Oxford: WileyBlackwell.

James, S. & Jacobsson, E. (u.å.). *Register-RCT – från spännande nyhet till etablerad metod?* Uppsala Clinical Research Center (UCR). <http://www.kvalitetsregister.se/download/18.72ba5c0e151e80f7a091f0e9/1451916518931/Register-RCT-fran-spännande-nyhet-till-etablerad-metod.pdf> [2016.05.13]

Jernberg, T. & Janzon, M. (2015). Sanningen om den kliniska verkligheten finns i registren. Registerbaserad randomiserad klinisk prövning kan ge representativ bild. *Läkartidningen*, 112:DLFX.

Jonasson, J.M., Ljung, R., Talbäck, M., Haglund, B., Gudbjörnsdottir, S. & Steineck, G. (2009). Insulin glargine use and short-term incidence of malignancies – a population-based follow-up study in Sweden. *Diabetologia*, 52.

- Knaapen, L. (2013). Being evidence-based in the absence of evidence: The management of non-evidence in guideline development. *Social Studies of Science* 43:5.
- La Folie, P. (2002). Epidemiologi vs klinisk prövning En match i fem ronder om vetenskaplig tillförlitlighet. *Läkartidningen*, 99.
- Lagerqvist, B. (u.å.) *RRCT Enorm potential, men hur får vi till det?* UCR, Kardiologen, Akademiska Sjukhuset, Uppsala. [http://qrcstockholm.se/wp-content/uploads/2015/11/Kurs\\_registerforskning\\_B\\_Lagerqvist\\_2015\\_nov.pdf](http://qrcstockholm.se/wp-content/uploads/2015/11/Kurs_registerforskning_B_Lagerqvist_2015_nov.pdf) [2016.05.13]
- Lee, S. J., Craig, C. E. & Weeks, J. C. (2000). Outcomes Research in Oncology: History, Conceptual Framework, and Trends in the Literature. *Journal of the National Cancer Institute*, 92:3.
- Lee, W.C., Dekoven, M., Bouchard, J., Massoudi, M. & Langer J. (2014). Improved real-world glycaemic outcomes with liraglutid versus other incretin-based therapies in type 2 diabetes. *Diabetes, Obesity and Metabolism* 16.
- Li, Q., Chitnis, A., Hammer, M. & Langer, J. (2014). Real-world clinical and economic outcomes of liraglutide versus sitagliptin in patients with type 2 diabetes mellitus in the United States. *Diabetes Therapy*, 5.
- Norlin, J., Schmitt-Egenolf, M., Steen Carlson, K. & Persson, U. (2014). Registerstudier mäter effektivitet av läkemedel i klinisk praxis. Exemplet PsoReg – omfördelning av systematisk terapi kan ge ökad effektivitet. *Läkartidningen*, 111:CTIE.
- Pocock, S.J. & Elbourne, D. R. (2000). Randomized Trials or observational tribulations? *The New England Journal of Medicine* ss, 342:25.
- Raskin, I. E., & Maklan, C. W. (1991). Medical treatment effectiveness research A View From Inside the Agency for Health Care Policy and Research. *Evaluation and the Health Professions* 14.
- Relman, A.S. (1988). Assessment and accountability. The third revolution in medical care. *The New England Journal of Medicine* 319.
- Risstad, H., Sovik, T.T., Engström, M., Aasheim, E.T., Fagerland, M.W., Fagevik Olsén, M., Kristinsson, J.A., le Roux, C.W., Bohmer, T., Birkeland, K.I., Mala, T. & Olbers, T. (2015). Five-Year Outcomes After Laparoscopic Gastric Bypass and Laparoscopic Duodenal Switch in Patients With Body Mass Index of 50 to 60. A randomized Clinical Trial. *JAMA Surgery*, 150:4.
- Rosati, R.A., Wallace, A.G. & Stead, E.A. (1973). The way of the future. *Arch Intern Med* vol 131.
- Rosén, M., Axelsson, S. & Lindblom, J. (2008). Släng inte ut observationsstudier med badvattnet. Bedöm deras kvalitet istället. *Läkartidningen* 105:46.
- Sager, M. & Bohlin, I. (2011). Avslutning. Evidensens öde avgörs i detaljerna. I Bohlin, I. & Sager, M. (red). *Evidensens många ansikten*. Lund: Arkiv förlag.
- Salive, M.E., Mayfield, J.A. & Weissman, N.W. (1990). Patient outcomes research teams and the Agency for Health care policy and research. *Health services research* 25:5.
- SBU (2012). *Hur bedömer SBU evidensens samlade styrka?* Statens beredning för medicinsk utvärdering (SBU). <http://www.sbu.se/sv/publikationer/vetenskap--praxis/vetenskap-och-praxis/hur-bedomer-sbu-evidensens-samlade-styrka/> [2016-05-12]



SBU (2014a). *Mall för kvalitetsgranskning av randomiserade studier*. Statens beredning för medicinsk utvärdering (SBU).

SBU (2014b). *Utvärdering av metoder i hälso- och sjukvården. En handbok*. 2 uppl. Stockholm: Statens beredning för medicinsk utvärdering (SBU).  
<http://www.sbu.se/globalassets/ebm/metodbok/sbushandbok.pdf> [2016-04-11]

SBU (2016). *SBU:s ordlista*. <http://www.sbu.se/sv/var-metod/sbu-ordlista> [2016-04-12]

SFS 2010:659. *Patientsäkerhetslag*. Stockholm: Socialdepartementet.  
[https://www.riksdagen.se/sv/DokumentLagar/Lagar/Svenskforfattningssamling/Patientsakerhetslag-2010659\\_sfs-2010-659/](https://www.riksdagen.se/sv/DokumentLagar/Lagar/Svenskforfattningssamling/Patientsakerhetslag-2010659_sfs-2010-659/) [2016-03-10]

Sismondo, S. (2010). *An Introduction to Science and Technology Studies*, 2ed. Singapore: Blackwell Publishing Ltd.

SKL (u.å.). *Förbättra vården med kvalitetsregister*.  
<http://www.kvalitetsregister.se/forbattravarden.373.html> [2016-04-11]

SKL (u.å.). *Stora framgångar för Diabetesregistret bara början*.  
<http://www.kvalitetsregister.se/forbattravarden/godaexempel/godaexempel/storaframgangarfordiabetesregistretbaraborjan.1405.htm> [2016-04-11]

Socialstyrelsen (2015). *Nationella riktlinjer för diabetesvård. Vetenskapligt underlag. Bilaga*. Socialstyrelsen, ss. 102-112. <https://www.socialstyrelsen.se/SiteCollectionDocuments/nr-diabetes-vetenskapligt-underlag.pdf> [2016-05-10]

Socialstyrelsen (u.å.). Register. [www.socialstyrelsen.se/register](http://www.socialstyrelsen.se/register) [2016-05-10]

Streptomycin in Tuberculosis Trials Committee (1948). Streptomycin treatment of pulmonary tuberculosis a medical research council investigation. *British Medical Journal*, 2:769.

Taube, A. (2008). Registerstudie endast ett steg i en vetenskaplig process. Genererar hypoteser snarare än prövar dem. *Läkartidningen*, 105:8.

Thörn, Å. & Järhult, B. (2015). Efter Riksrevisionens kritik: Dags för förnyad diskussion om kvalitetsregistrens roll. *Läkartidningen*, 112:DFDC.  
<http://www.lakartidningen.se/Opinion/Debatt/2015/04/Dags-for-fornyad-diskussion-om-kvalitetsregistrens-roll/> [2016-05-13]

Toresson, M. (2015). RRCT kan vara framtiden för registerforskningen. *LäkemedelsVärlden. (LmV)*, 9 november. <http://www.lakemedelsvarlden.se/nyheter/rrct-kan-vara-framtiden-registerforskningen-14748> [2016-05-13]

Walach, H., Falkenberg, T., Fonnebo, V., Lewith, G. och Jonas, W.B. (2006). Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BioMed Central Medical Research Methodology*, 6:29.

Wennberg, J. E. (1984). Dealing with medical practice variations: a proposal for action. *Health Affairs*, 3:2:7-8, 15, 21, 24. doi: 10.1377/hlthaff.3.2.6.

Zethelius, B., Gudbjörnsdottir, S., Eliasson, B., Eeg-Olofsson, K. & Cederholm, J. (2014). Level of physical activity associated with risk of cardiovascular diseases and mortality in patients with type-2 diabetes: report from the Swedish National Diabetes Register. *European Journal of Preventive Cardiology*, 21:2